JN

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification:** **C07H 21/04, C07K 14/46, C07K 16/28, C12P 19/34, C12Q 1/68**	**A1**	(11) **International Publication Number:**	**WO 00/18787**
		(43) **International Publication Date:**	06 April 2000 (06.04.2000)

(21) **International Application Number:** PCT/US99/22429

(22) **International Filing Date:** 28 September 1999 (28.09.1999)

(30) **Priority Data:**
60/102,031 28 September 1998 (28.09.1998) US

(60) **Parent Application or Grant**
WASHINGTON UNIVERSITY [/]; (). PERMUTT, M., Alan [/]; (). INOUE, Hiroshi [/]; (). MUECKLER, Mike [/]; (). PERMUTT, M., Alan [/]; (). INOUE, Hiroshi [/]; (). MUECKLER, Mike [/]; (). REED, Janet, E. ; ().

Published

(54) **Title: GENE MUTATED IN WOLFRAM SYNDROME**
(54) **Titre: MUTATION DE GENE ASSOCIEE AU SYNDROME DE WOLFRAM**

(57) Abstract

This invention provides a novel gene, WFS1, isolated from human chromosome 4p. Mutation of the WFS1 gene is associated with the development of Wolfram Syndrome. The WFS1 gene, along with cDNAs, encoded protein and antibodies immunologically specific for the protein, provide a biological marker for early diagnosis of the syndrome, and for predicting predisposition of an individual for the syndrome. The gene also will be useful in gene replacement therapy, or for development of new methods and agents for treating Wolfram Syndrome.

(57) Abrégé

L'invention se rapporte à un nouveau gène, WFS1, que l'on a isolé sur le chromosome humain 4p. La mutation de ce gène WFS1 est associée au développement du syndrome de Wolfram. La protéine codée par ce gène WFS1, associé à des séquences d'ADNc, ainsi que des anticorps spécifiques de ladite protéine d'un point de vue immunologique fournissent un marqueur biologique permettant un diagnostic précoce du syndrome ainsi qu'un diagnostic de la prédisposition d'un sujet à développer ce syndrome. Ce gène peut également être utilisé en thérapie génique de remplacement ou pour le développement de nouvelles méthodes et de nouveaux agents permettant de traiter le syndrome de Wolfram.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07H 21/04, C07K 14/46, 16/28, C12Q 1/68, C12P 19/34 — A1

(11) International Publication Number: WO 00/18787

(43) International Publication Date: 6 April 2000 (06.04.00)

(21) International Application Number: PCT/US99/22429

(22) International Filing Date: 28 September 1999 (28.09.99)

(30) Priority Data:
60/102,031 28 September 1998 (28.09.98) US

(71) Applicant: WASHINGTON UNIVERSITY [US/US]; Campus Box 8013, 600 South Euclid Avenue, St. Louis, MO 63110 (US).

(71)(72) Applicants and Inventors: PERMUTT, M., Alan [US/US]; 6341 Washington Avenue, St. Louis, MO 63130 (US). INOUE, Hiroshi [JP/JP]; Yamaguchi University School of Medicine, Third Dept. of Internal Medicine, Ube Yamaguchi (JP). MUECKLER, Mike [US/US]; Washington University School of Medicine, Dept. of Cell Biology and Dept. of Surgery, Division of Human Molecular Genetics, St. Louis, MO 63110 (US).

(74) Agents: REED, Janet, E. et al.; Dann, Dorfman, Herrell and Skillman, 1601 Market Street, Suite 720, Philadelphia, PA 19103 (US).

(81) Designated States: AU, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENE MUTATED IN WOLFRAM SYNDROME

(57) Abstract

This invention provides a novel gene, *WFS1*, isolated from human chromosome 4p. Mutation of the *WFS1* gene is associated with the development of Wolfram Syndrome. The *WFS1* gene, along with cDNAs, encoded protein and antibodies immunologically specific for the protein, provide a biological marker for early diagnosis of the syndrome, and for predicting predisposition of an individual for the syndrome. The gene also will be useful in gene replacement therapy, or for development of new methods and agents for treating Wolfram Syndrome.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Description

GENE MUTATED IN WOLFRAM SYNDROME

FIELD OF THE INVENTION

This invention relates to the field of diagnosis and treatment of certain forms of diabetes. In particular, this invention provides a novel gene, the disruption of which is associated with Wolfram Syndrome, and methods of using the gene and specific mutants thereof as diagnostic tool for prediction and early detection of the syndrome.

BACKGROUND OF THE INVENTION

Various scientific and scholarly articles are referred to in brackets throughout the specification. These articles are incorporated by reference herein to describe the state of the art to which this invention pertains.

Wolfram syndrome (WFS)(OMIM 222300) was first described in 1938 as a combination of familial juvenile-onset diabetes mellitus and optic atrophy. Other clinical features subsequently emerged and accordingly, WFS is also referred to as the DIDMOAD (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness) syndrome. Most patients with this

progressive disorder eventually develop all four cardinal manifestations, and die prematurely with widespread atrophic changes throughout the brain. Insulin-requiring diabetes mellitus occurs with mean age of onset at 6-8 years. When examined, pancreatic islets display atrophic and insulin-producing β-cells selectively absent. The disease is believed to account for 1/150 patients with young-onset insulin-requiring diabetes mellitus.

The pathogenesis of Wolfram syndrome is unknown. Diagnosis is usually made in offspring of unaffected often-related parents, suggesting autosomal recessive inheritance. Linkage of the gene to markers on chromosome 4p (Polymeropoulos et al., Nature Genetics 8: 95-97, 1994) was reported, and later confirmed (Collier et al., Am L Hum Genet 59: 855-863, 1996). Recombinants placed the gene in an interval between markers 5.5 cM apart.

Isolation and characterization of the gene or genes associated with Wolfram Syndrome is vital for the prediction, diagnosis and, ultimately, treatment of the disease. Currently there is no way of knowing for sure if an individual is predisposed to Wolfram Syndrome, particularly children too young to have developed the characteristic symptoms. Early diagnosis may lead to effective methods of treatment. Identification and isolation of the gene or genes associated with Wolfram Syndrome would further enable the development of screening procedures to assist in genetic counseling, as well as enabling detailed study of its function and subsequent development of therapeutic methods or agents.

SUMMARY OF THE INVENTION

In accordance with the present invention, a gene is provided, whose mutation is highly correlated with Wolfram Syndrome in humans. The symptoms of Wolfram

Syndrome consist of diabetes insipidus, diabetes mellitus, optic atrophy and deafness. This gene is found in human genome between markers *D4S500* and *D4S431* on chromosome 4p, and is referred to herein as *WFS1*. One variant of *WFS1* is found in SEQ ID NO:1.

Another aspect of the invention comprises several isolated nucleic acids from the human and mouse *WFS1* genes. In a preferred embodiment, these nucleic acids are the genomic sequence of the *WFS1* gene from human (SEQ ID NO:1), a cDNA sequence from the human *WFS1* gene (SEQ ID NO:2), and a cDNA from the mouse *WFS1* gene (SEQ ID NO:4). In another preferred embodiment, the isolated nucleic acids are substantially the same or 60% homologous to SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:4. In yet another preferred embodiment, the nucleic acids encode SEQ ID NO:3 and SEQ ID NO:5, substantially the same variants of SEQ ID NO:3 and SEQ ID NO:5, variants with at least 60% homology to SEQ ID NO:3 and SEQ ID NO:5, and variants of SEQ ID NO:3 with the mutations and polymorphisms listed in Table 1. In a final preferred embodiment, the isolated nucleic acids are oligonucleotides (SEQ ID Nos:6-41) that have been designed based on SEQ ID NO:1.

In accordance with another aspect of the invention, a selection of isolated polypeptides is provided, which result from the expression of part or all of the human and mouse *WFS1* genes. In one preferred embodiment, the polypeptides are encoded by SEQ ID NO:3 and SEQ ID NO:5. In another preferred embodiment, the polypeptides are substantially the same as SEQ ID NO:3 and SEQ ID NO:5, or at least 60% homologous to SEQ ID NO:3 and SEQ ID NO:5.

In accordance with another aspect of the invention, a selection of antibodies that are immunologically specific to the aforementioned

polypeptides is provided.

In accordance with another aspect of the invention, a series of methods is provided, which use the *WFS1* gene to genetically characterize mammalian subjects. In a preferred embodiment, the *WFS1* gene is used to make labeled probes which are used to detect *WFS1* nucleic acids. In another preferred embodiment, variant forms of *WFS1* are sequenced, compared to the sequence of *WFS1*, and mutations and polymorphisms determined. In a particularly preferred embodiment, human genomic DNA is sequenced and compared to SEQ ID NO:1. In a very particularly preferred embodiment, the primers in Table 2 are used to sequence the variant human gene. In yet another preferred embodiment, restriction enzymes are selected that differentially digest the wild type and variant gene, *WFS1* nucleic acids are digested and separated by size, and the *WFS1* nucleic acids are detected. In a particularly preferred embodiment, the human DNA is digested, and the primers in Table 2 are used to amplify the DNA before digestion.

Other features and advantages of the present invention will be better understood by reference to the drawings, detailed descriptions and examples that follow.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Pedigrees of Wolfram syndrome families, with individuals designated by disease status (bold-lined symbols affected, thin-lined symbols unaffected), and with derived haplotypes of chromosome 4p markers. Bold, underlined numbers represent disease-associated chromosomes. Italicized, underlined numbers refer to disease-associated chromosomes with historical recombinants. Consanguineous Japanese families: Fig. 1a, WS-1; Fig. 1b, WS-2; Fig. 1c,

WS-3 (an affected daughter, not shown, was deceased before initiation of the study). Caucasian families: Fig. 1D, WS-4; and Fig. 1E, WS-5. Markers were ordered from telomeric to centromeric as described in Example 1.

Figure 2. Physical map of the Wolfram syndrome critical region. Fig. 2A: The horizontal line at the top of the figure represents a portion of chromosome 4p, with the centromere to the right and pter to the left. The dashed line represents the interval of *D4S500-D4S431*, the critical region of the Wolfram syndrome gene. P1 and BAC clones are represented as lines. Their length reflects the number of STSs and not the actual size. The name of each clone is given to the right of the line. Marker names are noted above the line and correspond to the symbols on the line. Fig. 2B: An expanded schematic of the genomic structure of the Wolfram gene, *WFS1*, with exons indicated by boxes. The entire gene is encompassed within a 33.4 kb region.

Figure 3. Expression of *WFS1* mRNA in adult tissues. Fig. 3A: Northern analysis with human adult polyA+ RNA (5 μg) derived from various tissues was hybridized with a ^{32}P-labeled 854-bp genomic fragment of the *WFS1* gene. Fig. 3B: Re-hybridization of the blot with ^{32}P-labeled β-actin cDNA as a loading control. Fig. 3C: Northern analysis with human adult total RNA (20 μg) hybridized with ^{32}P-labeled *WFS1* cDNA. Fig. 3D: Re-hybridization of the blot in c. with ^{32}P-labeled ribosomal cDNA as a loading control.

Figure 4. Hydrophobicity analysis was conducted by the method of Kyte and Doolittle (J Molec Biol 157: 105-132, 1982) using a window size of 9 amino acid residues. Average hydrophobicity values are plotted as a function of position along the polypeptide chain.

Figure 5. Comparison of human and mouse *WFS1*

protein sequences. Plain text indicate identical residues. Amino acid gaps between human and mouse proteins are shown by dashes. The locations of the mutations found in Wolfram syndrome patients, 3 missense, the premature stop codon (X), 2 deletions, and the 7 bp repeat insertion, are indicated below the sequences. A predicted prenyltransferase α-subunit repeat structure (A493 to L502) is underlined.

Figure 6. Co-segregation of *WFS1* mutations with the disease phenotype in Wolfram syndrome families. Gig. 6a: The 1685del(N) in family WS-2, showing that each affected child (III-1, -2, -3, and -4) is homozygous for the mutation. Fig. 6b: Sequence chromatograms of the region of exon 8 showing the 15 bp deletion in a patient homozygous for the mutation, along with a normal control. Fig. 6c: Segregation of the 1681C to T and the microscopic deletion in family WS-5. The 1681C to T mutation destroys a BsmF1 site in a 766 bp PCR (polymerase chain reaction) fragment. The hemizygous T(-) affected children (II-1, -2, and -4) have only a 766 bp uncut fragment, while the hemizygous C(-) mother (I-4) and unaffected daughter (II-3) have 686 bp and 80 bp bands, and the heterozygous CT father (I-3) has 766 bp, 686 bp and 80 bp bands. The symbols "-" and "+" refer to the absence or presence of enzyme.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Various terms relating to the biological molecules of the present invention are used hereinabove and also throughout the specification and claims.

With reference to nucleic acids molecules, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that

is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

With respect to RNA molecules, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to proteins or peptides, the term "isolated protein (or peptide)" or "isolated and purified protein (or peptide)" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the

compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the similar sequences of the nucleic or amino acids thus define the differences. In the comparisons made in the present invention, the BLAST programs (NCBI) and parameters used therein were employed, and the DNAstar system (Madison, WI) was used to align sequence fragments of genomic DNA sequences. However, equivalent alignments and similarity/identity assessments can be obtained through the use of any standard alignment software. For instance, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the default parameters used (gap creation penalty=12, gap extension penalty=4) by that program may also be used to compare sequence identity and similarity.

The term "substantially the same" refers to nucleic acid or amino acid sequences having sequence variation that do not materially affect the nature of the protein (i.e. the structure, stability characteristics, substrate specificity and/or biological activity of the protein). With particular reference to nucleic acid sequences, the term "substantially the same" is intended to refer to the coding region and to conserved sequences governing expression, and refers primarily to degenerate codons encoding the same amino acid, or alternate codons encoding conservative substitute amino acids in the encoded polypeptide. With reference to amino acid sequences, the term "substantially the same" refers generally to conservative substitutions and/or variations in regions of the polypeptide not involved in

determination of structure or function.

The terms "percent identical" and "percent similar" are also used herein in comparisons among amino acid and nucleic acid sequences. When referring to amino acid sequences, "percent identical" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical amino acids in the compared amino acid sequence by a sequence analysis program. "Percent similar" refers to the percent of the amino acids of the subject amino acid sequence that have been matched to identical or conserved amino acids. Conserved amino acids are those which differ in structure but are similar in physical properties such that the exchange of one for another would not appreciably change the tertiary structure of the resulting protein. Conservative substitutions are defined in Taylor (1986, J. Theor. Biol. 119:205). When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence analysis program.

With respect to antibodies, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to oligonucleotides or other single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally

used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the

transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid, phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA". Such a nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

The term "selectable marker gene" refers to a gene encoding a product that, when expressed, confers a selectable phenotype such as antibiotic resistance on a transformed cell.

The term "reporter gene" refers to a gene that encodes a product which is easily detectable by standard methods, either directly or indirectly.

A "heterologous" region of a nucleic acid construct is an identifiable segment (or segments) of the nucleic acid molecule within a larger molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA

that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, a heterologous region is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein. The term "DNA construct", as defined above, is also used to refer to a heterologous region, particularly one constructed for use in transformation of a cell.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Description

In accordance with the present invention, a novel human gene, *WFS1*, has been isolated. The inheritance of a mutated *WFS1* gene is highly correlated

with the development of Wolfram Syndrome in the human population. Normal function of *WFS1* appears to be essential for survival of pancreatic islet β-cells and neurons. Included in the invention is the method for using the gene sequence to genetically screen for presence of the potentially mutated forms of the gene for diagnosis and prognosis of the disease in patients.

The *WFS1* gene in humans spans 33.4 kb on chromosome 4p and is composed of 8 exons (Fig. 2B; SEQ ID NO:1). The ~5kb of sequence upstream of the start of *WFS1* open reading frame is expected to contain one or more transcriptional or translational regulatory elements. The *WFS1* human cDNA is 3.688 kb long (SEQ ID NO:2) and encodes a predicted protein 890 residues long with a predicted molecular mass of 100.29 kDa (SEQ ID NO:3). Comparison of the *WFS1* cDNA sequence with those in public databases found no related genes. A mouse cDNA of *WFS1* has also been isolated. The mouse cDNA (SEQ ID NO:4) is 3511 nucleotides long and has 83.9% nucleotide identity to the human gene. The predicted protein sequence encoded by the mouse cDNA (SEQ ID NO:5) has a 86.1% amino acid similarity to the predicted human *WFS1* protein.

The inventors have isolated the *WFS1* gene represented by SEQ ID NO:1 from the human genome by positional cloning. Previous work had isolated the Wolfram gene between markers *D4S432* and *D4S431* on chromosome 4p, 5.5 cM (~5500 kb) apart. In the development of this invention, five families with individuals having typical Wolfram syndrome phenotypes were genotyped with genetic markers shown to locate between *D4S432* and *D4S431* by physical and/or radiation hybrid mapping (see Example 1). The region containing the Wolfram gene was thus narrowed further to a region

between *D4S500* and *D4S431*. The critical region between *D4S500* and *D4S431* was estimated to be <250 kb as determined by contig mapping of BAC and P1 genomic clones. Three clones were sequenced and the sequence of much of the contig region was determined.

Exon trapping of two BAC clones was employed to generate expressed sequence tags (ESTs) of the region and then to determine areas with open reading frames that would be likely to contain a gene. Among a number of ESTs isolated was one predicted to be the 3' end 1.8 kb exon of a gene. A genomic fragment of this region hybridized to a 3.7 kb RNA on a Northern blot. The gene was determined to be expressed in all tissues, but surprisingly was most abundant in pancreatic islets compared to that in the exocrine pancreas. The abundance of *WFS1* RNA in pancreatic islets was particularly revealing because one of the outcomes Wolfram Syndrome is atrophy of pancreatic islets.

A full length cDNA clone of *WFS1* was obtained by screening a human infant brain cDNA library (SEQ ID NO:2). This clone was 3,688 nucleotides long and contained an appropriate start methionine, open reading frame, and polyadenylation signal. Comparison of the cDNA sequence of *WFS1* with those in public databases revealed no related genes. Translation of the cDNA sequence predicts a polypeptide of 890 amino acid residues with a molecular mass of 100.29 kDa. The protein is distinguished grossly by the presence of 3 structural domains, a hydrophilic N-terminal region of ~300 residues, a hydrophilic C-terminal region of ~240 residues, and a central hydrophobic core of ~350 residues. Inspection of the hydrophobicity curve suggests the presence of ~10 transmembrane segments.

When patients with Wolfram Syndrome were checked for mutagenesis in the *WFS1* gene, seven mutations

were found in the full-length clones derived from the original EST. PCR (polymerase chain reaction) was used to amplify exons of *WFS1* from subjects with Wolfram Syndrome, and the products were sequenced. Comparison of these sequences with the wild type gene revealed probable loss of function mutations in all cases, as described in greater detail below.

The following description set forth the general procedures involved in practicing the present invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit the invention. Unless otherwise specified, general cloning procedures, such as those set forth in Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989) (herein "Sambrook et al.") or Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1999) (herein "Ausubel et al.") are used. Unless otherwise specified, general genome analysis procedures were used, such as set forth in Genome Analysis: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1997).

I. Preparation of *WFS1* nucleic acid molecules, encoded proteins and immunologically specific antibodies

A. Nucleic Acid Molecules

Nucleic acid molecules comprising part or all of the *WFS1* gene of the invention may be prepared by two general methods: (1) they may be synthesized from appropriate nucleotide triphosphates, or (2) they may be isolated from biological sources. Both methods utilize protocols well known in the art.

The availability of nucleotide sequence information, such as Sequence I.D. Nos. 1, 2 and 4, enables preparation of an isolated nucleic acid molecule

of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramadite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct may be purified by high performance liquid chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current oligonucleotide synthetic methods. Thus, for example, a double-stranded DNA molecule several kilobases in length may be synthesized as multiple smaller segments of appropriate complementarily. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an appropriate vector.

WFS1 nucleic acid sequences may be isolated from appropriate biological sources using methods known in the art. In a preferred embodiment, a human genomic clone is isolated from a human genomic P1 library. In another preferred embodiment, a cDNA clone is isolated from the Marathon-Ready human fetal brain λ gt10 cDNA library (Clontech). In yet another preferred embodiment, a mouse cDNA is isolated from a mouse pancreatic β-cell line (MIN6) cDNA library. The isolation of human and mouse clones is not limited to the aforementioned libraries, and other commercially available human and mouse libraries may be used. Alternatively, cDNA or genomic clones from other species may be obtained.

In accordance with the present invention, nucleic acids having the appropriate sequence homology

with part or all of Sequence I.D. Nos. 1, 2 or 4 may be identified by using hybridization and washing condition of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., using a hybridization solution comprising: 5 x SSC, 5 x Denhardt's reagent, 1.0% SDS, 100 μg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hour. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2 x SSC and 1% SDS; (2) 15 minutes at room temperature in 2 x SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1 x SSC and 1% SDS; (4) 2 hours at 42-65°in 1 X SSC and 1% SDS, changing the solution every 30 minutes. In a preferred embodiment, hybridizations are performed in hybridization solution comprising 0.5 M $NaPO_4$, 2 mM EDTA, 7% SDS and 0.1% sodium pyrophosphate (pH 7.1) at about 65°C for 20 hours. For high-stringency conditions, membranes are subsequently washed sequentially for 1 hour each in: (1) 2X SSC, 0.5X SET, 0.1% sodium pyrophosphate; and (2) 0.1X SSC, 0.5X SET, 0.1% sodium pyrophosphate. For low-stringency conditions, membranes are washed at 50°C for 30 minutes in 2X SSC, 0.5X SET, 0.1% sodium pyrophosphate.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology (Sambrook et al., 1989):

$$T_m = 81.5°C + 16.6\text{Log}\,[Na+] + 0.41(\%\ G+C) - 0.63\ (\%\ \text{formamide}) - 600/\#\text{bp in duplex}$$

As an illustration of the above formula, using [N+] = [0.368] and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1%

decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

The stringency of the hybridization and wash depend primarily on the salt concentration and temperature of the solutions. In general, to maximize the rate of annealing of the probe with its target, the hybridization is usually carried out at salt and temperature conditions that are 20-25°C below the calculated T_m of the of the hybrid. Wash conditions should be as stringent as possible for the degree of identity of the probe for the target. In general, wash conditions are selected to be approximately 12-20°C below the T_m of the hybrid. In regards to the nucleic acids of the current invention, a moderate stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured salmon sperm DNA at 42°C, and wash in 2X SSC and 0.5% SDS at 55°C for 15 minutes. A high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured salmon sperm DNA at 42°C, and wash in 1X SSC and 0.5% SDS at 65°C for 15 minutes. A very high stringency hybridization is defined as hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured salmon sperm DNA at 42°C, and wash in 0.1X SSC and 0.5% SDS at 65°C for 15 minutes.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, genomic clones are maintained in a COS-7 cells in the vector pSPL3 (Life Technologies, Inc.). In another preferred embodiment, PCR products are subcloned into pAMP10 using the UDG cloning kit (GIBCO

BRL), and propagated in a suitable *E. coli* host cell. In another preferred embodiment, clones are maintained in plasmid cloning/expression vector, such as pGEMT (PROMEGA), and propagated in *E. coli*.

WFS1 nucleic acid molecules of the invention (including those containing known polymorphisms and mutations) include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides (sense or antisense strands of DNA or RNA) having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of Sequence I.D. Nos. 1, 2 and 4. Such oligonucleotides are useful as probes for detecting *WFS1* genes (and specific mutations) in test samples, e.g. by PCR amplification, or as potential regulators of gene expression.

B. Proteins and Antibodies

A full-length *WFS1*-encoded protein of the present invention may be prepared in a variety of ways, according to known methods. The protein may be purified from appropriate sources, e.g., human or animal cultured cells or tissues, by immunoaffinity purification. However, due to the limited amount of such a protein that may be present in a sample at any given time, particularly in tumors or tumor cell lines, conventional purification techniques are not preferred in the present invention.

The availability of the isolated *WFS1* coding sequence enables production of protein using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such a pSP64 or pSP65 for *in vitro*

transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocytes. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or BRL, Rockville, Maryland.

Alternatively, the recombinant protein may be produced by expression in a suitable procaryotic or eukaryotic system. For example, part or all of a DNA molecule, such as the cDNA having SEQ ID NO:2 or No. 4, may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*, or into a baculovirus vector for expression in an insect cell. Such vectors comprise the regulatory elements necessary for expression of the DNA in the bacterial host cell, positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The protein produced by *WFS1* gene expression in a recombinant procaryotic or eukaryotic system may be purified according to methods known in the art. A commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted from the host cell, to be easily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein. Such methods are commonly used by skilled practitioners.

Proteins prepared by the aforementioned methods may be analyzed according to standard procedures. For

example, such proteins may be subjected to amino acid sequence analysis, according to known methods.

Included in the present invention are antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward *WFS1*-encoded proteins may be prepared according to standard methods. Monoclonal antibodies may be prepared, which react immunospecifically with various epitopes of the proteins. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies that immunospecifically interact with *WFS1*-encoded proteins can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of antibodies are described below.

II. Uses of *WFS1* Nucleic Acids, Encoded Proteins and Immunologically Specific Antibodies

A. *WFS1* Nucleic Acids

Nucleic acids comprising part or all of the *WFS1* gene may be used for a variety of purposes in accordance with the present invention. As illustrated in Example 1, selected *WFS1* sequences (DNA, RNA or fragments thereof) may be used as probes to identify mutations or rearrangements in a patient's DNA, and/or monitor the level of *WFS1* transcripts in tissues. As discussed earlier, *WFS1* mutations are associated with the occurrence of Wolfram Syndrome, a disease of autosomal recessive inheritance. Early identification of patients

destined to develop Wolfram Syndrome may lead to preventive therapies. Identification of heterozygous individuals that are at risk of having a child with Wolfram Syndrome will be very useful in genetic counseling.

WFS1 sequences may be utilized as probes in a variety of assays known in the art, including but not limited to: (1) *in situ* hybridization; (2) Southern hybridization; (3) northern hybridization; and (4) assorted amplification reactions, such as polymerase chain reaction (PCR). In a preferred embodiment, large deletion and premature termination mutations are detected by separation on acrylamide or agarose gel electrophoresis and Southern blotting with probes made from *WFS1* gene sequences. Knowledge of the wildtype sequence allows the identification of point mutations in non-functional *WFS1* genes. In another preferred embodiment, mutated genes are differentiated from wildtype genes by using restriction enzyme sites that appear or disappear as the result of the mutation. *WFS1* nucleic acids are digested and the fragments are then separated and probed as described above.

Both of the above-mentioned preferred embodiments are illustrated in Example 1. The human genomic *WFS1* sequence was used to design screening procedures to quickly screen all the individuals in a Wolfram Syndrome patient's extended family. In the case of large deletion mutations, the screen entailed amplifying the region of the gene encompassing the deletion, then separating the products by agarose gel electrophoresis and performing Southern blot hybridization using a labelled wild type PCR fragment. In the case of point mutations, the PCR primers were designed as above, the mutant and wildtype products were digested with a restriction enzyme that was specific to

either the sequence at the point mutation or the corresponding wildtype sequence. The products of the digestion, along with undigested control nucleic acids, were separated and detected as above. In all of the six extended families studied, inheritance of a homozygous complement of a mutated *WFS1* gene was consistent with the development of the disease and the pedigree of the family. Of particular interest are the three mis-sense mutations in the human *WFS1* predicted polypeptide, whose wild type sequence is found to be conserved in the predicted mouse polypeptide. This sequence conservation, together with the mutational effect of non-conservative amino acid substitutions at these sites, suggests that these amino acids are critical for gene function. A primary screen of these sequences is a useful way to expedite the identification of mutations. Other critical mis-sense mutations may also be useful in connection with this invention. These other mutations can be found using procedures detailed in Example 1 and others well known in the art.

The *WFS1* nucleic acids of the invention may also be utilized as probes to identify related genes either from humans or from other species. In a preferred embodiment, a cDNA has been isolated from mouse insulinoma cDNA library by standard screening methods and RACE PCR. The mouse cDNA was 3,511 nucleotides long with 83.9% nucleotide identity to the coding sequence of the human gene and 86.1% predicted amino acid similarity. While the cloning of the mouse cDNA is illustrated in Example 1, those skilled in the art will appreciate that the isolation of *WFS1* from other species is by no means limited to mouse. Other mammalian species of interest include, but are not limited to, cow, cat, dog, horse, pig and rat. As is well known in the art, hybridization stringency may be adjusted so as to allow hybridization

of nucleic acid probes with complementary sequencing of varying degrees of homology.

The cDNA from the mouse *wfs1* gene is very useful because it allows further study of the gene and Wolfram Syndrome in system more conducive to research than human. The mouse clone may be used to create a targeting construct, which can be used for the targeted mutagenesis of the endogenous mouse gene. By creating both null and site-directed mutants, a mouse model system for Wolfram Syndrome can be generated. This mouse system can subsequently be used to elucidate the cellular function of the Wolfram gene product, as well as assist in developing therapies for the syndrome. Systems other than mouse can also be used to advantage. These systems include, but are not limited to animal models developed in mouse, various cultured human and mammalian cell systems (e.g., mouse and rat insulinoma cells) and frog oocyte expression systems.

As described above, the coding region of *WFS1* may also used to advantage to produce substantially pure *WFS1* encoded proteins or selected portions thereof. As described below, these proteins may also be used in diagnosis and therapy of Wolfram Syndrome.

B. Proteins and Antibodies

The *WFS1*-encoded protein, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies, which also may serve as sensitive detection reagents for the presence and accumulation of the *WFS1*-encoded polypeptide in cultured cells or tissues from living patients (the term "patient" refers to both humans and animals). Because the *WFS1*-encoded protein has not yet been isolated from natural sources, such antibodies will greatly accelerate the identification, isolation and characterization of this protein in mammalian cells and

tissues. Recombinant techniques enable expression of fusion proteins containing part or all of the *WFS1*-encoded protein. The full-length protein or fragments of the protein may be used to advantage to generate an array of monoclonal antibodies specific for various epitopes of the protein, thereby potentially providing even greater sensitivity for detection of the protein in cells or tissues. Monoclonal antibodies specific to variant portions of the *WFS1* polypeptide may also be used to advantage in diagnosing presence of a variant form of the gene.

Polyclonal or monoclonal antibodies immunologically specific for the *WFS1*-encoded protein may be used in a variety of assays designed to localized and/or quantitate the protein. Such assays include, but are not limited to: (1) flow cytometric analysis; (2) immunochemical localization of the protein in cultured cells or tissues; and (3) immunoblot analysis (e.g., dot blot, Western blot) of extracts from cells and tissues. Additionally, as described above, such antibodies can be used for the purification of *WFS1*-encoded proteins (e.g., affinity column purification, immunoprecipitation).

The following example is provided to describe the invention in greater detail. It is intended to illustrate, not to limit, the invention.

EXAMPLE 1
Isolation and Identification of Wolfram Gene and Mutants

METHODS

Patients and families. Three families originated from Japan: WS-1 (Nanko et al., Brit J Psychiatry 161: 282, 1992), WS-2 (Higashi, Am J Otology

12: 57-60, 1991, and WS-3 (Maruta et al., Clin Neurol 27: 725,732, 1987). Cell lines for the Caucasian family WS-4 were obtained from the NIGMS Human Genetic Mutant Cell Repository, Camden, NJ. (family #1157). Family WS-5 is Caucasian Australian and WS-6 Saudi Arabian. Minimum criteria for diagnosis were young-onset insulin-dependent diabetes mellitus and progressive optic atrophy.

Microsatellite analysis. Initial genotyping was carried out as described previously (Nestorowicz et al., Hum Molec Genet 7: 1119-1128, 1998; Inoue et al., Diabetes 45: 789-794, 1996) with markers reported in the 1996 Genethon Microsatellite Map (*D4S127, Hox7, D4S412, D4S3023, D4S2925, D4S431, D4S2935, D4S3007, D4S394, D4S2983, D4S2923*). Second genotyping was performed with markers (*D4S2957, D4S2375, A348XA5, D4S827, D4S500, and D4S2366*), shown to locate in the interval of *D4S3023-D4S431* by physical and radiation hybrid mapping projects of chromosome 4. The STANFORD CHR 4 YAC MAP project and the CHROMOSOME 4 SUMMARY MAP were viewed at WEB sites (http://shgc.stanford.edu/, http://cedar.genetics.soton.ac.uk/). Primer sequences were obtained from the Genome Database (GDB - http://www.hgmp. mrc.ac.ok/gdb/gdbtop.html).

P1/BAC library screening. A human genomic P1 library (HD-K) was screened for clones containing *D4S500* and *D4S431* by PCR, using primers developed to specifically hybridize with those markers. Sequence-tagged sites (STSs) (SP6 and T7) from P1s 102C5, 89C1 and 77B6, as well as *D4S500* and *D4S431*, were used for BAC library screening by either PCR or by direct hybridization of library grid blots (Research Genetics, Inc., Huntsville, Alabama).

Exon-trapping, sample and shotgun megabase genomic sequencing. Restriction fragments from BAC 460K9 and 33H22 were cloned into the BamHI site of pSPL3,

transfected into COS-7 cells and spliced products obtained by RT-PCR were subcloned into pAMP10 using the UDG cloning kit (GIBCO BRL) and sequenced.

Sample and shotgun sequencing of BAC460K9 and 33H22 was accomplished as described (Wilson and Mardis, in *Genome Analysis: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, NY, 1997). Similarity searches with known genes and ESTs were performed using BLAST programs (NCBI). The sequencing project was compiled using LaserGene software (DNAStar, Madison, WI). Sequencing was performed on an ABI 373 with Prism dye terminator kits.

Northern blot analysis, cDNA screening, and 5' RACE analysis. A multiple adult human tissue polyA+ RNA Northern blot (Clontech- MTN human I) was probed with an 854-bp fragment (nt 2133-2986 *WFS1* cDNA) as described by the manufacturer. Total cellular RNA was isolated (Chomczynski and Sacchi, Anal Biochem 162: 156-159, 1987) and Northern analysis performed with 20 μg of RNA by hybridization with *WFS1* cDNA labeled with [α-^{32}P]dCTP as previously described (Ferrer et al., Diabetes 46: 386-392, 1997). RNA quality and loading was checked by staining the gel for ribosomal RNA and by hybridization with either β-actin or ribosomal RNA.

The same 854-bp fragment was used for screening an infant brain λ gt10 cDNA library. Six overlapping clones containing the *WFS1* gene were isolated. 5' RACE analysis was performed using Marathon-Ready human fetal brain cDNA (Clontech, Palo Alto, CA). The RACE products were subcloned into pGEMT vector (Promega Biotech, Madison WI) and sequenced. A mouse pancreatic β-cell line (MIN6) cDNA library was screened with human *WFS1* cDNA using standard reduced stringency conditions, and a 3.0 kb clone was isolated. A primer corresponding to the

predicted 5'-untranslated region, based on the mouse EST sequence (GenBank AA021827 and AA692227) that was homologous to the human cDNA, was synthesized and used for RT-PCR with MIN6 RNA (5'-CGGTTTCGGAGCAACTTCGC-3' , SEQ ID NO:42 and 5'- CACCTCAGCCTCGTTCTCAG -3', SEQ ID NO:43).

Mutation detection. M13 universal primer sequence was incorporated into the 5' terminus of primers for direct sequence analysis using an ABI automated DNA sequencer Model 373 (Chadwick et al., Biotechniques 20: 676-683, 1996). Primers used for exon amplification, genomic sequencing and mutation detection are as in Table 2.

The 2812del(TC) was detected with a labeled PCR fragment on a denaturing Long Ranger polyacrylamide gel (FMC Bioproducts). The 1685del(N_{15}) was detected by electrophoresis of PCR fragments on 4% agarose gels. The 2341C to T was detected by PCR, digestion with EcoNI (New England Biolabs (NEB)) and polyacrylamide-gel electrophoresis. The mutated allele was observed as 209 bp and 35 bp fragments.

The 2254G to T mutation was detected by PCR, digestion with AvaII (NEB) and agarose gel electrophoresis, with two bands (240 bp, 137 bp) in the wild allele, and three (240 bp, 120 bp, 17 bp) in the mutant. The 2114G to A mutation was detected by PCR, digestion with Tfi I (NEB) and agarose gel electrophoresis. The wild allele has one band of 528 bp, and the mutated allele has two bands (396 bp, 132 bp). The 1681C to T mutation destroys a BsmF1 restriction site, and primers are set 8a (Table 2), with resulting fragments described in Fig. 6c.

GenBank Accession Numbers. The human *WFS1* and

mouse *wfs1* cDNA sequences are deposited in GenBank: #AF084481 and #AF084482.

RESULTS

Linkage Analysis. Linkage studies were conducted on three Japanese families (WS-1, -2, and -3) and two Caucasian families (WS-4, -5) (Figs. 1A-1E), each with at least two individuals having typical Wolfram syndrome phenotypes. When genotyping with chromosome 4p markers used previously (Polymeropoulos et al., Nature Genetics 8: 95-97, 1994; Collier et al., Am J Hum Genet 59: 855-863, 1996), estimates for recombination fractions (θ) between WFS and the markers confirmed close linkage (lod =3.99 for *D4S431* at θ =0.05). Multipoint analysis with GENEHUNTER (version 1.1) (Kruglyak et al., Am J Hum Genet 59: 1347-1363, 1996) gave a lod >6.0 for the region encompassed by markers *D4S827 and D4S394*.

Haplotype Analysis and Mapping by Recombination. Haplotypes were constructed by inspection. The boundaries for the WFS gene had been defined by a telomeric recombinant at *D4S432* (Collier et al., 1996), and centromeric recombinants at *D4S431* (Polymeropoulos et al., 1994; Collier et al., 1996). Recombinants in families WS-1 - WS-6 were identified by genotyping with genetic markers shown to locate between *D4S432* and *D4S431* by physical and/or radiation hybrid mapping. In family WS-1, subject III-2, recombination of the telomeric region from *D4S827* was observed (Fig. 1A). In the Japanese family WS-2 (Fig. 1B), all affected subjects (III-1, -2, -3, and -4) were haploidentical centromeric to *D4S500*, suggesting the presence of a historical recombinant in the unrelated father (II-5). The centromeric boundary was confirmed as *D4S431* by a recombinant in WS-5, subject III-2 (Fig. 1E). We concluded that the WFS gene likely lies within the

interval *D4S500-D4S431*.

Construction of a genomic contig across the WFS critical region. In the STANFORD CHR 4 YAC MAP project, *D4S827* and *D4S431* are within two overlapping YACs encompassing the region (204H9 and 420E4). Further, *D4S500* and *D4S431* were located on the same YAC (420E4). A contig was constructed with genomic human P1/BAC clones, and a contiguous genomic map encompassing the WFS gene was obtained (Figure 2). STS content mapping confirmed overlapping clones. As the interval between *D4S500* and *D4S431* was covered by a P1 (102C5) and a BAC clone (33H22), the critical region was estimated to be <250 kb.

Identification of candidate genes within the *WFS* region, and cloning the *WFS1* gene. By exon trapping of BACs 33H22 and 460K9, an expressed sequence tag (EST) was found that resulted in the cloning of the γ-isoform of the B regulatory subunit of the human protein phosphatase 2A (PP2ABRγ). The genomic structure of PP2ABRγ was determined, and direct sequence analysis of probands excluded this gene. In parallel, we initiated large-scale genomic sequencing from BACs 460K9 and 33H22, and P1 102C5. A total of ~180 kb of sequence was analyzed. Among a number of EST matches, one was predicted to be the 3'-end exon containing a 1.8 kb open reading frame (exon 8, Fig. 2B). This EST was further evaluated as it was unambiguously within the critical region.

The size of the full-length mRNA and pattern of tissue expression was determined by PCR amplification of an 854-bp genomic fragment of the region that was hybridized to a multiple-tissue Northern blot with polyA+ RNA. A major transcript of ~3.7 kb was expressed in all tissues including pancreas (Fig. 3A). Northern analysis of total RNA (20 μg) revealed the gene most abundantly

expressed in pancreatic islets compared to that in exocrine pancreas (Fig. 3C).

A full-length clone was obtained by screening a human infant brain cDNA library. Six clones were isolated and subsequently 5' RACE analysis was performed. These analyses yielded a composite cDNA sequence of 3.688 kb. The longest open reading frame extended from nt 171 to 2843. The methionine at position 171 was chosen as the translation initiation codon primarily because it conforms to Kozak's rule (Kozak, Mamm Genome 7: 563-574, 1996). A consensus polyadenylation site (aataaa) was located at position 3615-20, 19 bases upstream from the polyA tail. The gene was named *WFS1*.

Predicted characteristics of the WFS1 protein, and cloning of the mouse cDNA. Comparison of the cDNA sequence of *WFS1* with those in public databases revealed no related genes. Translation of the cDNA sequence predicts a polypeptide of 890 amino acid residues with a molecular mass of 100.29 kDa. Hydrophobicity analysis of the deduced amino acid sequence is presented in Figure 4. The protein is distinguished grossly by the presence of 3 structural domains, a hydrophilic N-terminal region of ~300 residues, a hydrophilic C-terminal region of ~240 residues, and a central hydrophobic core of ~350 residues. Inspection of the hydrophobicity curve suggests the presence of ~10 transmembrane segments, if it is assumed that this region of the protein consists of α-helical segments. Comparison of the predicted amino acid sequence with entries in the Prosite database produced a single match to the prenyltransferase α-subunit repeat structure.

A mouse *wfs1* cDNA was isolated from a mouse insulinoma (MIN6) (Ishihara et al., Diabetologia 36: 1139-1145) cDNA library, and completed by RT-PCR. The mouse *wfs1* cDNA was 3511 nucleotides with 83.9%

nucleotide identity to the coding sequence of the human gene, and 86.1% amino acid similarity (Figure 5).

Genomic structure of the WFS1 gene and mutations in WFS patients. The genomic structure of *WFS1* was determined by comparison of cDNA and genomic sequences obtained by shotgun sequencing of BAC460K9 and 33H22 and sequences in the Stanford Human Genome Center database (http://www.shgc.stanford.edu). The gene was found to be composed of eight exons (Fig. 2B) in 33.4 kb of genomic DNA.

For mutation screening, exons were amplified and sequenced from patients' genomic DNA. A TC deletion at position 2812 for subject WS-1 III-2 predicted a frameshift at codon 882, designated del882fs/ter937 (Table 1), with absence of the normal stop codon at 891 and the introduction of a new downstream termination codon. The predicted *WFS1* protein contains 937 amino acids, 47 more than the normal protein. All 3 affected sibs (WS-1 III-1, -2, and -4) were homozygous for this mutation, while the unaffected sib and the parents were heterozygous, indicating a disease-specific mutation. The 2812delTC mutation was not found in 80 healthy control Japanese subjects (160 chromosomes, see Table 1).

In other WFS families, six additional mutations were found in exon 8 (Table 1). In family WS-2, affected offspring (III-1, -2, -3, and -4) inherited a 15 bp deletion resulting in del508YVYLL, homozygous by descent from related heterozygous parents. Co-segregation of this deletion with the WFS phenotype is shown in Fig 6A. A sequence chromatogram from an affected child homozygous for the 15 bp deletion is shown in Fig. 6B. In family WS-3, both affected offspring (II-1 and -2) were homozygous for a 2341 C to T transversion resulting in a P724L mutation. In the Caucasian family WS-4, all affected offspring (II-1, -3, and -4) were found to be compound heterozygotes for a 2254 G to T transition

resulting in a G695V (paternal) mutation, and a 2114 G to A transversion resulting in a W648X (maternal) mutation. The W648X mutation predicts a premature termination, and lack of 242 amino acids of the C- terminus. In each of these 4 families the mutations were shown to co-segregate with the disease phenotype, both by sequencing and by either size change, or alteration of a restriction endonuclease site. None of the mutations were found in Japanese or Caucasian control subjects. No other coding variants were found on sequencing the entire gene in each proband.

In the Australian family WS-5, a 1681C to T transversion (P504L) was observed. This mutation destroys a BsmF1 restriction site. In Fig. 6C, the father (I-3) is shown to be heterozygous for 1681C to T, and the mother (I-4) is homozygous 1681C. Yet surprisingly all affected offspring (II-1, -2, and -4) appeared to be homozygous 1681T. The unaffected child (II-3) appeared to be homozygous for 1681C. The most likely explanation for these findings is that the mother's chromosome, inherited by each child (see haplotypes in Fig. 1E), harbored a microscopic deletion for the *WFS1* gene, and that the affected offspring were hemizygous for the P504L mutation.

A sixth family (WS-6) with one 10 year old affected child and two apparently unaffected younger sisters became available for analysis. The parents were Saudi Arabian first cousins. The affected child was homozygous for a 7bp repeat insertion at 1610 (CTGAAGG), resulting in a predicted frame shift and premature termination of the protein at codon 544. The parents were heterozygous, while the unaffected sisters were heterozygous and homozygous normal respectively. Sequence analysis also revealed a number of silent and intronic variants (polymorphisms) in various families (see Table 1).

Table 1. Mutations and polymorphisms in *WFS1*

Mutation	Amino acid	Exon/intron	Family	Control chromosomes
2812del(TC)	del882fs/ter937	Exon 8	WS-1	160[a]
1685del(CCTGCT CTATGTCTA)	del508YVYLL	Exon 8	WS-2	160 [a]
2341C to T	P724L	Exon 8	WS-3	160 [a]
2254G to T	G695V	Exon 8	WS-4	160[b]
2114G to A	W648X	Exon 8	WS-4	160[b]
1681 C to T	P504L	Exon 8	WS-5	160[b]
del*WFS1*[c]	---	---	WS-5	ND
1610insCTGAAGG	ins483fs/ter544	Exon 8	WS-6	160[d]
Changes of uncertain effect				
1167G to A	I333V	Exon 8	WS-4 (same chr. as W648X)	ND
Polymorphisms				
854G to C	D268D	Exon 6	WS-4	ND
1355T to C	V395V	Exon 8	WS-4	ND
1457C to T	C429C	Exon 8	WS-3	ND
1537A to G	R456H	Exon 8	WS-1,-2,-3,-4, normals	ND
1545C to T	L459L	Exon 8	WS-3	ND
1570T to C	L507L	Exon 8	WS-4	ND
1815C to T	L549L	Exon 8	WS-4	ND
1925C to T	F585F	Exon 8	WS-4	ND
2002G to A	H611R	Exon 8	WS-1,-2,-3, normals	ND
2603A to G	K811K	Exon 8	WS-4	ND
2735G to A	S855S	Exon 8	WS-4	ND
1032-5C to G		Intron 7	WS-1,-2,-3,-4	ND

a-Japanese
b-Caucasian
c-Not confirmed
d-Palestinian Arabs

Table 2. Primers for amplification, sequencing and mutation detection (SEQ ID NOS: 6-41, consecutively).

DNA Fragment	Primers	Product Size
Exon 1*	5'TGTAAAACGACGGCCAGTCTCGTGCAGAAGGCCGCGCT3' 5'CAGGAAACAGCTATGACCGCCCACAGCCACCGCGCCAC3'	247 bp
Exon 2	5'TGTAAAACGACGGCCAGTCTGTCTCCAGCAGACACTAA3' 5'CAGGAAACAGCTATGACCCACAATGCTGAACTGCAGAG3'	276 bp
Exon 3	5'TGTAAAACGACGGCCAGTCTGAAGACCCTCATGCCTTG3' 5'CAGGAAACAGCTATGACCACACTTCTCTGTGGGCTGTG3'	276 bp
Exon 4	5'TGTAAAACGACGGCCAGTTCGGAGAATCTGGAGGCTGA3' 5'CAGGAAACAGCTATGACCCATTACAAGCTGCTCAACCC3'	253 bp
Exon 5	5'TGTAAAACGACGGCCAGTCGAAAGCCTTCCAGGCAGAG3' 5'CAGGAAACAGCTATGACCCTATGGGAAGGTCCTGGCTC3'	353 bp
Exon 6	5'TGTAAAACGACGGCCAGTCTAGGAACAGTGCGCCAGTT3' 5'CAGGAAACAGCTATGACCATGGAGTCGCACAGGAAGGA3'	268 bp
Exon 7	5'TGTAAAACGACGGCCAGTGCCCATGCTGTTTTCTCTCA3' 5'CAGGAAACAGCTATGACCCCGAGGACACATCCTTATGA3'	371 bp
Exon 8a	5'TGTAAAACGACGGCCAGTCCTCGTTCCCACGTACCATC3' 5'CAGGAAACAGCTATGACCGTAGCAGTAGGTGCCCTTGA3'	766 bp
Exon 8b	5'TGTAAAACGACGGCCAGTCCTGGTCGTCCTCAATGTCA3' 5'CAGGAAACAGCTATGACCCATAGAACCAGCAGAACAGC3'	503 bp
Exon 8c	5'TGTAAAACGACGGCCAGTTGGTTCACGTCTCTGGAGCT3' 5'CAGGAAACAGCTATGACCGAGTTGTAGACCTTCATGCC3'	240 bp
Exon 8d*	5'TGTAAAACGACGGCCAGTGGGCATGAAGGTCTACAACT3' 5'CAGGAAACAGCTATGACCGAACTTCTTGATGTGGCAGG3'	362 bp
Exon 8e	5'TGTAAAACGACGGCCAGTCTGGATGCGCTGCCTCTACG3' 5'CAGGAAACAGCTATGACCTCAGGCCGCCGACAGGAATG3'	523 bp
Exon 8f	5'TGTAAAACGACGGCCAGTTCGCCTTCGACTTCTTTTTC3' 5'CAGGAAACAGCTATGACCCCAAACAAATAAGAAATGCT3'	499 bp
2812del(TC)	5' GCC CAG CTC TCG CCC ACC AG 3' 5' TCA GGC CGC CGA CAG GAA TG 3'	120 bp
1685del(N)$_{15}$	5' CCT GGT CGT CCT CAA TGT CA 3' 5' GGT AGG GCA CAA GGT AGC AG 3'	119 bp
2341C to T	EcoNI-F;5' GGGCATGAAGGTCTACAACTCCA 3' EcoNI-R;5' CCGTAGAGGCAGCGCATCCAGTCGCCGAcctAGAAC3'**	244 bp
2254G to T	5'-GAGGGCATGAAGGTCTACAA-3' 5'- CCCACGGTAATCTCAAACTT-3'	377 bp
2114G to A	5'-TAGTGTGCCCCTGCTGTTGC-3' 5'-CCCACGGTAATCTCAAACTT-3'	528 bp

* Due to the inability to directly sequence these PCR products, the fragments were subcloned into pGEM-T Easy Vector (Promega) as described by the manufacturer and several colonies sequenced for each individual.

** Because originally there was no appropriate restriction enzyme site to distinguish mutated alleles, the EcoNI-R primer was modified (3-base change, tga to cct, underlined in primer sequence) and a new EcoNI site was introduced to the mutated allele (CC2341(C/T)NNNNNAGG).

DISCUSSION

Consanguineous families from isolated regions of Japan provided the genetic material that led to the discovery of mutations in *WFS1* in WFS patients of diverse genetic backgrounds. We believe that mutant alleles at *WFS1* are responsible for the disease for several reasons, beyond the fact that the gene maps to the critical region. In each of the six pedigrees, mutant alleles of *WFS1* co-segregated with the disease phenotype. *WFS1* was shown to be expressed in brain, pancreatic islets, and in a β-cell insulinoma cell line, consistent with the disease phenotype. Seven different mutations were found, as well as a presumed microscopic deletion. The three missense mutations were evolutionarily conserved between the mouse and human (Figure 5), further suggesting their biological significance. None of the mutations were observed in normal chromosomes.

The Australian Caucasian family WS-5 was particularly interesting, as each affected child appeared to be homozygous for a P504L mutant allele inherited from the heterozygous father (Fig. 6C). Repeat sampling and analysis confirmed these results. Analysis of the mother's DNA with new markers between *D4S500* and *D4S431* suggested that the deletion was confined to a region of <170 kb. Recently a patient with another autosomal recessive disorder was observed to be heterozygous for a missense mutation in combination with a partial deletion of a gene (Ries et al., Human Mutation 12: 44-51, 1998).

The expression pattern of *WFS1* appeared ubiquitous by Northern analysis of polyA+ RNA (Fig. 3A). Yet interestingly, the most prominent mRNA observed in total RNA was that in pancreatic islets (Fig. 3C). This high level of expression of *WFS1* in islets might explain why the earliest manifestation of WFS is insulin-deficient diabetes mellitus (Barrett and Bundey, J Med Genet 34: 838-841, 1997). Further analysis of the cell biology of *WFS1* will be accomplished through generation

of specific antibodies, monitoring expression in cultured cells, and gene targeting to define genotype/phenotype relationships.

Swift et al. hypothesized that heterozygous carriers of the gene for WFS were 26-fold more likely to require psychiatric hospitalization than non-carriers (Swift et al., Molecular Psychiatry 3: 86-91, 1998). Blackwood et al (Blackwood et al., Nature Genetics 12:427-430, 1996) reported highest lod scores with markers mapping to the region *D4S431-D4S403* in a genome scan of a large family with bipolar affective disorder. These findings suggest that mutations in *WFS1* might be implicated in patients with psychiatric diseases.

Mutations in *WFS1* appear to result in premature death of pancreatic islet β-cells leading to juvenile onset insulin-requiring diabetes mellitus (Karasik et al., Diabetes Care 12: 135-138, 1989). The β-cell damage in autoimmune Type I diabetes likely results from the interaction of the HLA locus as a major susceptibility gene, along with multiple minor gene defects. In contrast, islet β-cell loss in WFS is monogenic in origin. Importantly, the *WFS1* gene appears to play a major role in maintaining normal islet β-cell function, as mutations in this gene alone can result in loss of islet β-cells. Genome scans for both Type I and Type II diabetes mellitus have not implicated major genes in the 4p region. Yet since these are complex diseases, mutations in *WFS1* might play a minor role in these more common forms of diabetes. In addition, *WFS1* may represent a new therapeutic target for treatment and prevention of diabetes mellitus and for neurodegenerative disorders.

The present invention is not limited to the embodiments described and exemplified above, but is capable of variation and modification without departure from the scope of the appended claims.

Claims

What is claimed:

1. A recombinant DNA molecule comprising a vector into which is inserted a heterologous DNA segment from human chromosome 4p, the segment being located between markers *D4S500* and *D4S431*, the segment comprising a gene, mutations of which are associated with Wolfram Syndrome.

2. The recombinant DNA molecule of claim 1, wherein the gene is composed of exons that form an open reading frame having a sequence that encodes a polypeptide about 880 to 900 amino acids in length.

3. The recombinant DNA molecule of claim 2, wherein the open reading frame encodes an amino acid having greater than 60% identity with SEQ ID NO: 3 or SEQ ID NO:5.

4. The recombinant DNA molecule of claim 4, wherein said open reading frame comprises a sequence having greater than 60% homology with SEQ ID NO:2 or SEQ ID NO: 4.

5. The recombinant DNA molecule of claim 1, wherein the gene is composed of exons having sequences greater than 60% homologous with the sequences of the corresponding exons in SEQ ID NO:1.

6. An oligonucleotide between about 10 and 100 nucleotides in length, which specifically hybridizes with a portion of the recombinant DNA molecule of claim 1.

7. An isolated nucleic acid molecule having a sequence that is part or all of a sequence selected from the group consisting of:

a) SEQ ID NO:1;

b) a variant of SEQ ID NO:1 that is substantially the same as SEQ ID NO:1 within the exons of SEQ ID NO:1;

c) a sequence having at least homology 60% to SEQ ID NO:1 within the exons of SEQ ID NO:1;

d) SEQ ID NO:2;

e) a variant of SEQ ID NO:2 that is substantially the same as SEQ ID NO:2;

f) a sequence having at least homology 60% to SEQ ID NO:2;

g) a sequence encoding a polypeptide substantially the same as SEQ ID NO:3;

h) a sequence encoding a polypeptide at least 60% homologous to SEQ ID NO:3;

i) a sequence encoding a polypeptide substantially the same as SEQ ID NO:3, that additionally comprises one or more of the sequence variants set forth in Table 1.

j) SEQ ID NO:4;

k) a variant of SEQ ID NO:4 that is substantially the same as SEQ ID NO:4;

l) a sequence having at least homology 60% to SEQ ID NO:4;

m) a sequence encoding a polypeptide substantially the same as SEQ ID NO:5; and

n) a sequence encoding a polypeptide at least 60% homologous to SEQ ID NO:5.

8. An oligonucleotide between 10 and 100 bases in length, that specifically hybridizes with a portion of the nucleic acid molecule of claim 7.

9. A polypeptide, which is produced by the expression of the nucleic acid molecule of claim 7.

10. Antibodies immunologically specific for the polypeptide of claim 9.

11. A polypeptide produced by expression of an isolated nucleic acid molecule comprising part or all of an open reading frame of a gene located on human chromosome 4p between markers *D4S500* and *D4S431*, mutations of which are associated with Wolfram Syndrome.

12. The polypeptide of claim 11, which comprises a hydrophilic N-terminal region of about 300 amino acid residues, a hydrophilic C-terminal region of about 240 residues, and a central hydrophobic core of about 350 residues.

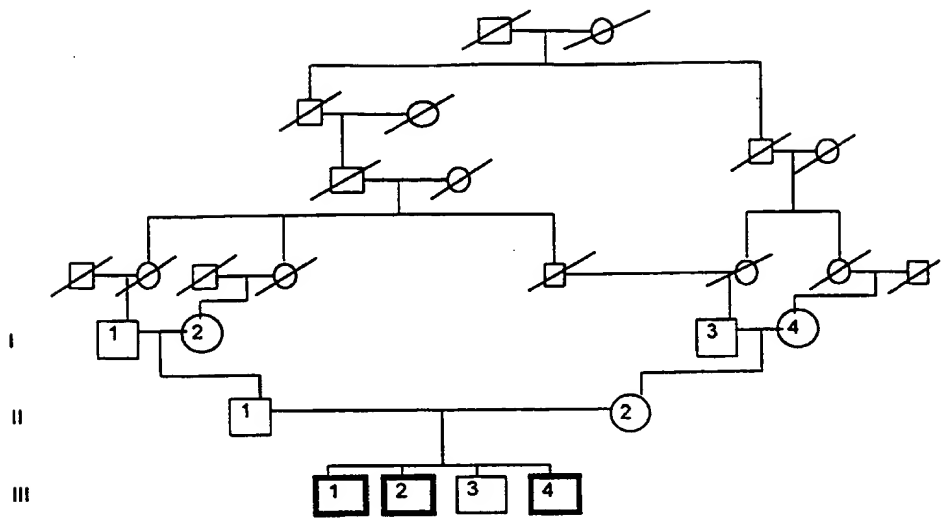
13. The polypeptide of claim 12, having an amino acid sequence substantially the same as part or all of SEQ ID NO: 3 or SEQ ID NO:5.

14. Antibodies immunologically specific for part or all of the polypeptide of claim 11.

15. A method for determining the predisposition of an individual to develop Wolfram Syndrome, which comprises examining a *WFS1* gene sequence of the individual for mutations resulting in expression of no gene product or a non-functional gene product, the mutations being indicative of the predisposition of the individual to develop Wolfram syndrome.

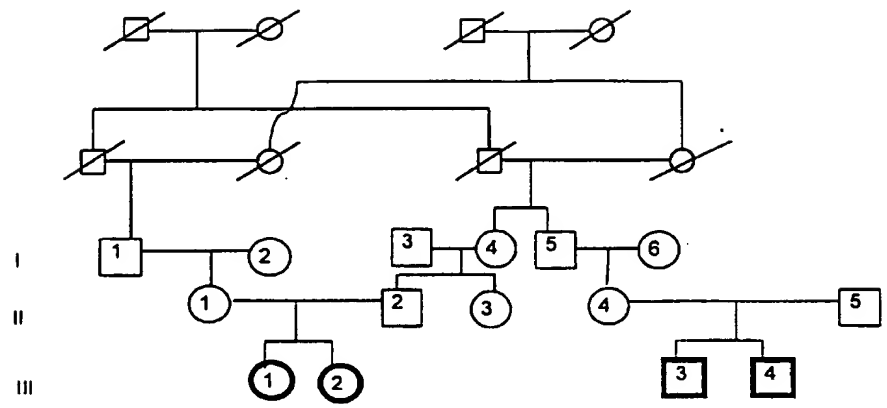
16. The method of claim 15, wherein the mutations comprises sequences selected from the group consisting of 2812del(TC), 1685del(CCTGCTCTATGTCTA), 2341C to T, 2254 G to T, 2114 G to A, 1681 C to T, and 1610ins(CTGAAGG).

17. The method of claim 16, wherein the mutations are detected by PCR amplification using primers selected from the group consisting of SEQ ID NOS: 32, 33, 34, 35, 36, 37, 38, 39, 40 and 41.



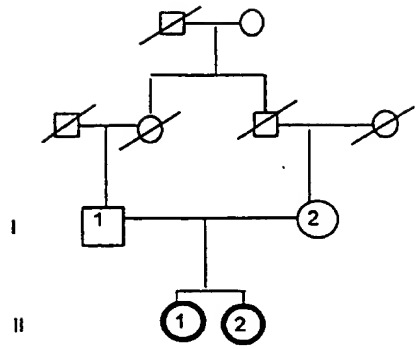
D4S127	3	4	1	4	3	4	3	3	3	4	4	4	3	3	3	4	3	4	1	4
D4S412	4	3	1	3	4	3	4	4	4	1	3	1	4	4	4	1	4	1	2	1
D4S3023	9	5	3	4	9	4	9	9	9	7	4	9	9	9	9	7	9	7	7	7
D4S2925	2	2	1	1	2	1	2	2	2	3	1	2	2	2	2	3	2	3	2	3
Hox7	3	3	3	1	3	1	3	3	3	1	1	3	3	3	3	1	3	1	3	1
D4S2957	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D4S2375	1	2	2	4	1	4	1	1	1	3	4	1	1	1	1	3	1	3	2	3
A348XA5	6	2	6	1	6	1	6	6	6	6	1	6	6	6	6	6	6	6	3	6
D4S827	6	2	1	3	6	3	6	6	6	3	3	6	6	6	6	3	6	3	1	3
D4S500	1	5	2	6	1	6	1	1	1	1	6	1	1	1	1	1	1	1	1	1
D4S431	3	6	4	6	3	6	3	3	3	3	6	0	3	3	3	6	3	6	4	6
D4S2360	2	3	2	5	2	5	2	2	2	2	5	2	2	2	2	3	2	3	5	3
D4S2935	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	3	4	3
D4S3007	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
D4S394	3	1	1	1	3	1	3	3	3	3	1	3	3	3	3	3	3	3	1	3
D4S2983	1	3	2	1	1	1	1	3	1	3	1	3	1	3	3	5	3	5	1	5
D4S2923	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 1a



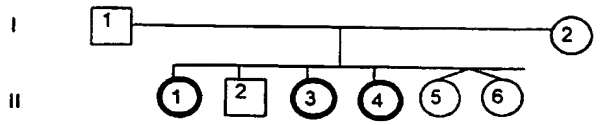
D4S127	1	4	1	1	1	1	1	4	1	4	1	4	3	1	1	1	1	1	5	1
D4S412	2	1	2	3	2	2	2	3	2	3	2	1	4	2	2	2	1	2	1	2	1
D4S3023	3	8	3	9	3	3	3	3	3	7	3	7	9	3	9	3	3	7	3	7	9	7
D4S2925	1	2	1	3	1	1	1	1	1	1	1	1	2	1	1	1	1	3	1	3	1	3
hox7	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	1	3	1
D4S2957	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D4S2375	4	2	4	5	4	4	4	4	4	4	4	2	1	4	4	4	4	4	4	4	1	4
A348XA5	2	2	2	2	2	2	2	2	2	2	2	2	6	2	6	2	2	3	2	3	2	3
D4S827	4	4	4	4	4	4	4	4	4	6	4	5	6	4	6	4	4	6	4	6	4	6
D4S500	2	9	2	2	2	2	2	2	2	4	2	3	1	2	2	2	2	11	2	11	A	11
D4S431	6	4	6	6	6	6	6	6	6	6	6	6	3	6	4	6	6	6	6	6	4	6
D4S2360	4	2	4	5	4	4	4	4	4	5	4	5	2	4	5	4	4	4	4	4	1	4
D4S2935	2	4	2	4	2	2	2	2	2	3	2	3	4	2	4	2	2	2	2	2	2	2
D4S3007	3	1	3	3	3	3	3	3	3	3	3	1	1	3	1	3	3	3	3	3	1	3
D4S394	1	3	1	5	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1
D4S2983	1	1	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1
D4S2923	3	2	3	1	3	3	3	3	3	1	3	1	1	3	1	3	3	3	3	3	1	3

Figure 1b



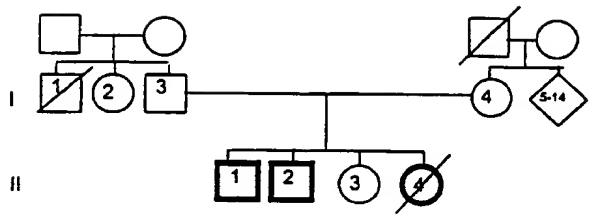
D4S127	4	1	1	1	1	1	3	1
D4S412	2	2	2	2	2	2	4	2
D4S3023	5	6	6	6	6	6	9	6
D4S2925	3	1	1	1	1	1	2	1
Hox7	3	3	3	3	3	3	3	3
D4S2957	1	1	1	1	1	1	1	1
D4S2375	2	4	4	4	4	4	1	4
A348XA5	4	2	2	2	2	2	6	2
D4S827	4	6	6	6	6	6	6	6
D4S500	6	6	6	6	6	6	1	6
D4S431	1	4	4	4	4	4	3	4
D4S2360	5	3	3	3	3	3	2	3
D4S2935	4	4	4	4	4	4	4	4
D4S3007	3	3	3	3	3	3	1	3
D4S394	4	1	1	1	1	1	3	1
D4S2983	1	5	5	5	5	1	1	5
D4S2923	1	1	1	1	1	1	1	1

Figure 1c



D4S127	2 4	2 2	4 4	2 2	2 2	4 4	4 2	2 4
D4S412	3 4	3 3	4 1	3 3	3 3	4 1	4 3	3 1
D4S3023	2 1	2 3	2 7	2 3	2 3	1 7	1 3	3 7
D4S2925	3 3	3 3	3 1	3 3	3 3	3 1	3 3	3 1
hox7	1 1	1 2	1 3	1 2	1 2	1 3	1 2	2 3
D4S2957	3 1	3 3	3 1	3 3	3 3	1 1	1 3	3 1
D4S2375	4 6	4 2	4 2	4 2	4 2	6 2	6 2	2 2
A348XA5	1 3	1 5	1 1	1 5	1 5	3 1	3 5	5 1
D4S827	7 6	7 7	7 6	7 7	7 7	6 6	6 7	7 6
D4S500	7 1	7 2	7 7	7 2	7 2	1 7	1 2	2 7
D4S431	5 4	5 4	5 5	5 4	5 4	4 5	4 4	4 5
D4S2360	1 1	1 5	1 2	1 5	1 5	1 2	1 5	5 2
D4S2935	2 4	2 4	2 2	2 4	2 4	4 2	4 4	4 2
D4S3007	3 3	3 4	3 3	3 4	3 4	3 3	3 4	4 3
D4S394	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1
D4S2983	4 3	4 5	4 1	4 5	4 5	3 1	3 5	5 1
D4S2923	1 1	1 3	1 1	1 3	1 3	1 1	1 3	3 1

Figure 1d



D4S2375	2 2	2 3	2 3	2 3	2 3	1 3
D4S500	1 1	1 2	1 2	1 2	1 2	2 2
D4S1054	1 2	1 1	1 1	2 1	1 1	1 1
JW3	1 4	1 2	1 2	4 2	1 2	3 2
D4S431	1 2	1 4	1 3	2 4	1 4	3 4
D4S3007	2 1	2 2	2 2	1 2	2 2	2 2

Figure 1e

Figure 2a

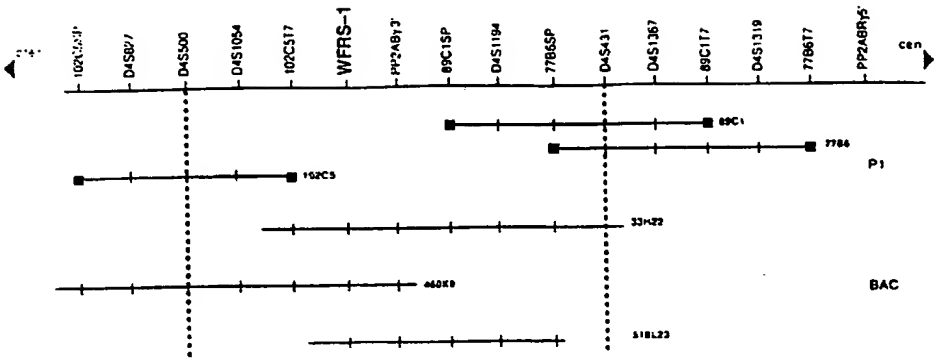


Figure 2b

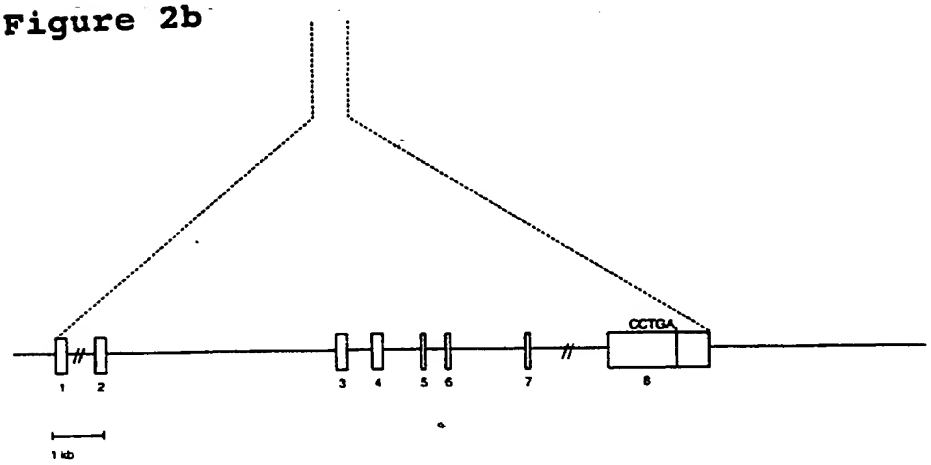


Figure 3a

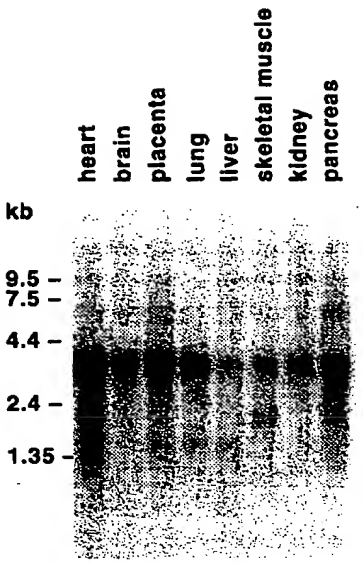


Figure 3b

Figure 3c

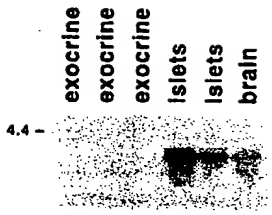


Figure 3d

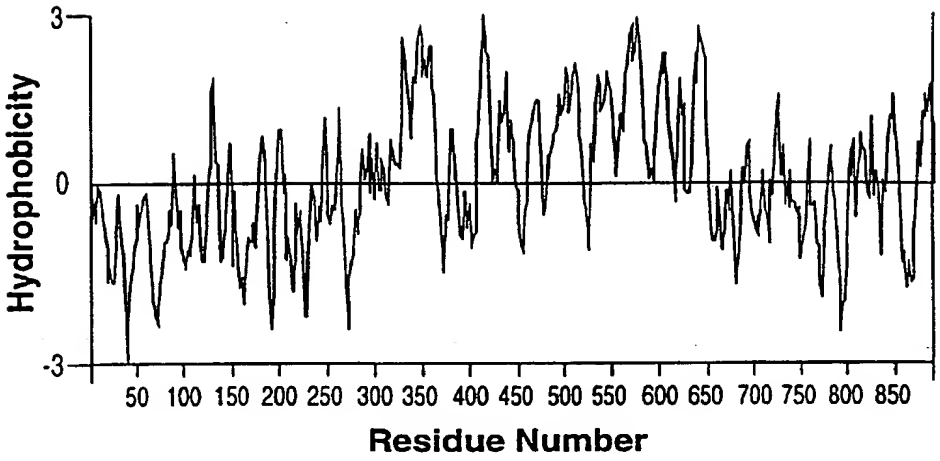

3
0
-3
Hydrophobicity
50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850
Residue Number

Figure 4

```
1    MNSGTPPPSPSGPPPPPAPQPQARARLNATASLEQDKIEPPRAPRPQADPSAGRSAGEAA Mouse
1    MDSNTAPLGPSCPQPPPAPQPQARSRLNATASLEQERSERPRAPGPQAGPGPGVRDAAAP Human

61   APEPRAPQTGSREETDRAGPMKADVEIPFEEVLEKAKAGDPKAQTEVGKHYLRLANDADE Mouse
61   A-EPQAQHTRSRERADGTGPTKGDMEIPFEEVLERAKAGDPKACTEVGKHYLQLAGDTDE Human

121  ELNSCSAVAWLILAAKQGRREAVKLLRRCLADRKGITSENEAEVKQLSSETDLERAVRKA Mouse
120  ELNSCTAVDWLVLAAKQGRREAVKLLRRCLADRRGITSENEREVROLSSETDLERAVRKA Human

181  ALVMYWKLNPKKKKQVAVSELLENVGQVNEQDGGVQPGPVPKSLQKaRRMLERLVSSESK Mouse
180  ALVMYWKLNPKKKKaVAVAELLENVGQVNEHDGGACPGPVPKSLQKGRRMLERLVSSESK Human

241  NYIALDDFVELTKKYAKGIIPTNLFLQDEDEDEDELAGKSPEDLPLRQKVVKYPLHAIME Mouse
240  NYIALDDFVEITKKYAKGVIPSSLFLQD-DEDDDELAGKSPEDLPLRLKVVKYPLHAIME Human

301  IKEYLIDVASKAGMHWLSTIVPTHHINALIFFFIISNLTIDFFAFFIPLVVFYLSFVSMV Mouse
299  IKEYLIDMASRAGMHWLSTIIPTHHINALIFFFIVSNLTIDFFAFFIPLVIFYLSFISMV Human

361  ICTLKVFQDSKAWENFRTLTDLLLRFEPNLDVEGAEVNFGWNHLEPYIHFLLSVVFVIFS Mouse
359  ICTLKVFQDSKAWENFRTLTDLLLRFEPNLDVECAEVNFGWNHLEPYAHFLLSVFFVIFS Human

421  FPLASKDCIPCSELAVISTFFTVTSYMSLSSSAEPYTRRALVTEVAAGLLSLLPTVPVDW Mouse
419  FPIASKDCIPCSELAVITGFFTVTSYLSLSTHAEPYTHRALATEVTAGLLSLLPSMPLNW Human

481  RFLKVLGQTFFTVPVGHFIILNVSLPCLLYVYLFYLFFRMACLRNFKGTYCYLVPYLVCF Mouse
479  PYLKVLGQTFITVPVGHLVVLNVSVPCLLYVYLLYLFFRMACLRNFKGTYCYLVPYLVCF Human
          Iins483fs/ter544          L504 Δ508

541  MWCELSVVILLQSTGLGLVRASIGYFLFLFALPILVAGLALMGTVQFARWFLSLDLTKIM Mouse
539  MWCELSVVILLESTGLGLLRASIGYFLFLFALPILVAGLALVGVLQFARWFTSLELTKIA Human

601  VTTVICGVPLLFRWWTKANFSVMGMVKSLTKSSMVKLILVWLTAILLFCWFYVYRSEGMK Mouse
599  VTVAVCSVPLLLHWWTKASFSVVGMVKSLTRSSMVKLILVWLTAIVLFCWFYVYRSEGMK Human
                                                        X648

661  VYNSTLTWQQYGFLCGPRAWKETNMARTQILCSHLEGHRVTWTGRFKYVRVTEIDNSAES Mouse
659  VYNSTLTWQQYGALCGPRAWKETNMARTQILCSHLGGHRVTWTGRFKYVRVTDIDNSAES Human
                                    V695

721  AINMLPFFLGDWMRCLYGEAYPSCSSGNTSTAEEELCRLKQLAKHPCHIKKFDRYKFEIT Mouse
719  AINMLPFFIGDWMRCLYGEAYPACSPGNTSTAEEELCRLKLLAKHPCHIKKFDRYKFEIT Human
          L724

781  VGMPF--GTNGNRGHEEDDITKDIVLRASSEFKDVLLNLRQGSLIEFSTILEGRLGSKWP Mouse
779  VGMPFSSGADGSRSREEDDVTKDIVLRASSEFKSVLLSLRQGSLIEFSTILEGRLGSKWP Human

839  VFELKAISCLNCMTQLSPARRHVKIEQDWRSTVHGALKFAFDFFFFPFLSAA Mouse
839  VFELKAISCLNCMAQLSPTRRHVKIEHDWRSTVHGAVKFAFDFFFFPFLSAA Human
                                                  Δ882fs/ter937
```

Figure 5

Figure 6a

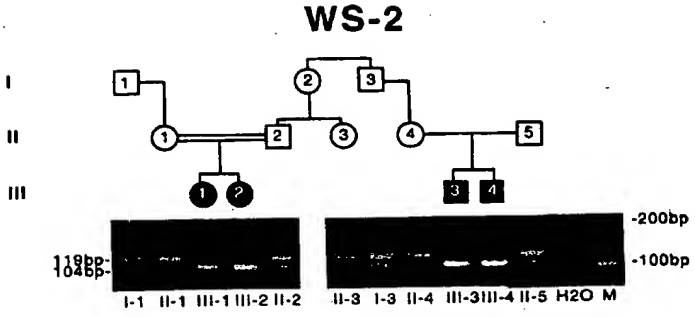


Figure 6b

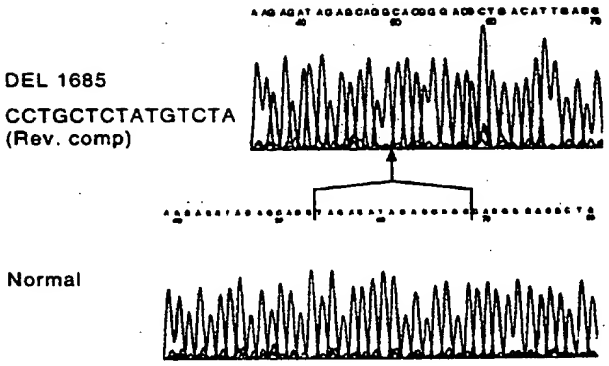
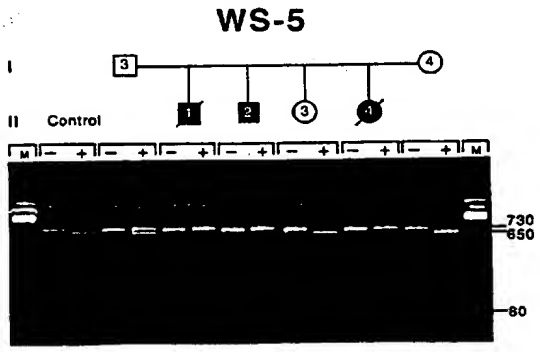


Figure 6c



SEQUENCE LISTING

<110> Permutt, M. Alan
Inoue, Hiroshi
Muekler, Mike

<120> Gene Mutated in Wolfram Syndrome

<130> WashU CI-0277

<150> US 60/102,031
<151> 1998-09-28

<160> 43

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 67212
<212> DNA
<213> Homo sapiens

<400> 1

aagcttgcta cactgcattt gattctgggc actgatagca gtacttggct gcagacactc 60
gttccaaaca ggccccattt ttccatctct gctgctgtta ttaggggagc ccttagactc 120
tcttgcagcg ccggaatagg cggtcaagac gtgtgttaat attgcaacag caaatataat 180
gaatctgcag ttggggacgc tgaggccgga gtggtggatg aaaggtggcc ggagcctttt 240
ccacgggtcc aaaccacctg ttacaggaga aggcgagcgg cctcgctaag caactggacg 300
ttccgcgggc ggggcggggc cggggcccga gtccgctcgg aaactttcgc tgtgggcgag 360
ccggacccgc cttctggccc ctcgggccca accacgcagg ggtagtgccg ggccgcagct 420
aggcgaggcg caccgtgatt ggcggagatg tggagtgatt ggcggctaca ccggccactc 480
agcaggccga gctggcgccg catccggggg gccgcgtctg gagtggaggg aggccgaagg 540
ccccgcccct gccccgcccc ctcgtgcaga aggccgcgct agccggctct tcagcagcga 600
gtgcagattg ctccccccgcg gccgcagatc tcccgtttgc gccgcgttca gctgctcccg 660
aacaactttt ctgccggccc agaggcccca gggcgtcgca gcgccgcgtg cggcccactc 720
acgggccggt gagtacttcg gcgctggggc agtggcgcgg tggctgtggg cagcgcgcag 780
gaggcgggca agagccctga ggcactgtcc tcctcgggcc tcagtttccc cttccgagct 840
gatgggtggc tggccccaaa gtcccgacaa ggtccccgaa gttggagggc cgggggtccc 900
gcccgctctg caacgcgcaa ggcgacccct gttcccgggc ccgaacgggt cacccggggg 960
gcgcgccccg gtcccccgcg cgctgtctgc atggagcccc gccgcgggcg ggacagcagg 1020
cccgagaggc ggagtggccg gcgggcaggc cccctccccg cgccccgcgc catccccccc 1080
gagttgcccg gtcatgcccc tcccccgcgg tcccgggacc cacggccgcc ttaggtgccc 1140
gccaggccgt tggaaggcct gggcccccgg gaggtgttgt gggggggaagg ggggttcagt 1200
cccccagggg gacccggccc gcccgagggg cagcttaggg gtcattctga atccccacag 1260
cccacttagg agcgctgtgg cagttcctct ccctctcggg gcccctcagt agagagttga 1320
ggggtttcta gtggcagcac ctggaccttg cctcgtatga actccacctt ctggatcatt 1380
agctttttgt ttgatcaacg aaaactttct ggagaatcct acccagccct tttggatgtg 1440
atggcgccct ggcctctcac ccacttagca gatgatgtta tgaattccaa ggaatcgcat 1500
ttgtgcaacc cccaccaccc agatttaacc ctctcagtag ccttaggcgg agaagttata 1560
gtccccgttt gacaccccag gaagctgagg ctcagcaggg tttggtgcca gcccaggacc 1620
acgaagctga tggggcagag ctggtgttag gcctctgccc tatctctcca gaagcagccc 1680
gtctctcctc accttggaca atgacccctg gcatttggcc caggctttgc agtttgtaaa 1740
ggctttctct tatgccagca tccatcactc caggagctgt gccctgggta ccccatgtgc 1800
tggctgcctg ttctctgcag caggagaggg agcccacggg aagcccagag gccacctttg 1860
tcactgtccc cacctgttta cttctgtgtc caagggtctg tgattcccat ggggtctccg 1920
ttgtatccaa ccagggattt ggagggaggc ttctccacat ttcagcctct tttccctacc 1980

tgctggtttg tcctcccttc gcagacacag gaaaagttgg ctgggagacc ccatctcacc 2040
ttctttccca ggatggaaga gcatggaggg cgtctctgac ttaggcctcc cctgcccatc 2100
ccggaccaat gctggacagg cttctggggc tggtcacaga ggggacctca gctcagtctc 2160
aacccccaga gatgtctgca ctgccagaca gacccgaaac acccatcctt tgacggtggc 2220
gataagaatc ataataggtc gggagcttta tgctacacgg gtcccctccc ctcccagttc 2280
cctacttccg gaagtggtac caccacagcc agccagcccc agccacaaac cagaccaccc 2340
cctgtgccct tcattcctca ctcccagttg ggtctctcct ggctcctgag cctcactcgg 2400
atgcctgccc tggttcagga ccctcgcctc tccctggacg acccccaggc tcctccctgg 2460
tcccgatccg ggcctggctg cacgcagccc tgctggttct tcagctgcca cttcccctctg 2520
ggctatcggg tcaggtccgc ctcccgagcc aggagccggc attctttcgg gccctcaggg 2580
acccgcctcg ctgctcactc agcctctccc tcctgtgtcc tctctccagc agtcacctcc 2640
cagccacctg cgtgttcctg gccacaagcc tggctcactg tcccctctgct gcctcctctt 2700
ctgggcacaa cccctgctgg aatgcccagc cacctcccag cttgcctggc tcctgcctgc 2760
ttgtccagaa ggctcagtca aggcctcctg caggatttct ggcctgcctc ttcccacagt 2820
ggagtgaggg ccccttctg tgccgtcttt gtgggtgtct gtttaggtgt taacccttcc 2880
catctagatg aagctgctcc aatgctggcc ctggacaaga cttctctagg gcaggggggc 2940
ccagcacagt gcttggcaca agatggacct accctcagtc cacacctgct gcccatccag 3000
gccgggcctg gagcaggtgc caggtgtgtg gtgggggctt ggccactatg gccgtcatga 3060
atgctgcttg gggcagtccc aattcaaacc cctctgttcc tgtgaatgtg ccacattgga 3120
tagagctgga atgttccttg gtttgtccct taatactctg agcaaaaata ggtacctggg 3180
atccatcaag ggctttgtgc ccttgtcttg aggagcagcc ctgtgccagg gggctgggtc 3240
aggccaggcc agagcagaca tcccagggcc aggacctcct gggcctgctc ccttagagct 3300
tgctcacaga tgggttttaa aataacattt gctttaattt caaaaacatc acacattcat 3360
tatagaaaac tagggaaata tgtttaaaag aagaataaaa tgaaaccgcg cataagcttt 3420
atccccagag acaacaagta tctttttttag ggggcgtttt cttccgtttt ttggcagata 3480
aattttgttt tcatcctaaa cagactccgg ctaatatctt ctgctgcttt aaatcctgcc 3540
ttaacagtgt taggcccacc ttttccttgt ctcaaaacgt tcatgtacga ggtcacgttt 3600
accagctgta cagcgttgca ccatatggag gggtgtaggc tttatagcca gccccttttct 3660
ggacagtgct gaaagtgcca tatgatggac attgtgtgtg agccccgagg gcatctccag 3720
tgactccect ggactggcat cctggaggca gaggcactga tcagagagag tgaacatcat 3780
cttggcttcc cacaggggct gtttctgagg cactggcttg ggagatgggg ttgccccccc 3840
agagagaaca gaatgggggt ggcctgggtg agggccagag cccttagggc agaaggccac 3900
gtgtggtcat tctgctgcgg gtctgagtgg atgcatggtg cctgcagccc tctccaggtg 3960
gcggcctctc agctgaactc cttggtgccg gtgtttgtgt ttggacggga taaaaatatg 4020
cctcattgct catgaggaaa gaatgctagc aactttggtt ttgtggatct tacacagtga 4080
cggcttctaa gccattgctg tggtcaccag tagggagttg gggctgagtg ccccaccgtg 4140
caggtgggga tctgagggtg cagagtggat gcagctggag atgcctggat gctggggtct 4200
gggggaggga cttgctaagg tcccacagca ttattgcctg gatggtgctc ctgccctagg 4260
ccaggctgcc tcaggttctg cccacatgca ttaagtgctt aacgtattcc tggcaaggca 4320
gccggtcttc cgagtggata agctgcagcc ctgccctcag ggctcacagt ggctgtaaca 4380
tgcaggaaga ctcagagagg gaaggaccag acagcaggtc cttccaccgg gagatctttt 4440
ctgatcttca ttgtccctgt gaccaaccat gaccaggaat tgttgtgctt gttttgaaat 4500
aacaagccct ttgatccaaa ccagacttgg ccaaccccca gggaacctga aaattaatcc 4560
tcctgtattc atccattctt actaaccaga gattttactt taaacagttt ttttttaact 4620
gcagtgtgac atacatggta cataggccag ctagaatctg ttttcctgaa attccttta 4680
tgaagtgttt tccaactgaa catgcctttg tttagcacct gggagtaatg gcttagggtt 4740
tttcttttcc tttttctttt ttttttttt gagacggagt ctcgctttgt cacccaggct 4800
ggagtgcagt ggcgcgatct cggctcactg caagctccat ctcctgggtt cacaccattc 4860
tcctgcctca gcctccccag tagttgggac tatcggcgcc cgccaccacg cctggctaat 4920
ttttttgtat ttttaataga gacggggttt cactgtgtta gccaggatgg tcttgatctc 4980
ctgaccttgt gatctgcccg cctcggcctc ccaaagtgct gggattacag gcgtgagcca 5040
ccgcgcctgg ccggcttagt tttttttaa tgcccaatag tttgctttct cttctgccac 5100
acacgtaatc cccatagcac ccctatgtgg gaggtgctat ccccatttta catatgggga 5160
aacagacaca gtgaggtcat gtggccacta gtaagggtaa ccgacgcccc acaccatgtt 5220
gcttccagct gtaacctcta gcagagggag ggctctgatg gagaggtgat cgcctgagtg 5280
ctgtgtgttt acctgagtag gtgaggaaga cagggactcg ggacagacat agcttggaag 5340
ctcccacaaa ggaggtatcg ctggagctct tgacttgaag ggtgcttagg gtctaggctg 5400
ggggaggggg gtaagaaggc attccgggaa gagggccctg cagaaacaga agtgtggagg 5460
tcagcatttg tatgctgatg ttgaagggta gccagtgttt tgcagtgaat gggccccggg 5520

agagagggtg cagtggcagt ctgatttcaa agtcatcttt ttcagaaatg agcttgcaga 5580
agcattagaa ttgggcaggt ctggatttga atttcaggtc tggtacctgg agctgtttga 5640
ccttgagcag gtccctcccc tccctgcacc tgggttcctg tctgtgaagt ggggtgaggt 5700
tgaaggataa aacatttccc catctataac tctgggatga caataggatc taccttccag 5760
gtccatgcga ggaactcacg acaataggac ctaccttcca gggccatacg agaaactcat 5820
gagaaaagac aagtacagag aacagagccg ttggtcttga tggcagagat gccctggcgt 5880
aggtcagccc tcagtaaagg tgccctctct ctctgtctga gaaagtgaca cctgacttcg 5940
gggcagtggg aaccatggga gagaaggtag acaaggccac gtgtcatttg aagacctgac 6000
ttttcaccat ctcaggtggt ccggtggcct gcaggaccgt ggttctgttt gcccacagcg 6060
tgtgggttct acgtgcgatc ctgtatttct ctgccacagc tgactaaggg gacacttagg 6120
tctgcacgtg cagttaaatg gctgcttttg tggaaggtgc cagttcccgg gagctgggaa 6180
cttggaaaca ggatggagca ggctcccaag cttgaccatg catggtttgc acctgggga 6240
tgcacccggg ggtgcttgac catgcatggt ttgcacctgg gggatgtttt acactaaggg 6300
ctcagtccct ctgagggagg aacgtgtgga ttgaaatctg cagagttctt gagcgacacc 6360
agtctgatct ttaggactca tttgcagcgg gaatggagca gaaattgttg tgtgtggagc 6420
ctggggccac tcctgactct tccagtcaca gctgccccca aaagtgacct ctgggcacct 6480
actcagacct ggctgctcag tgccaagcac tgggtgttca gaggtggcca ggggactccc 6540
tcgtgctgcc catggtgtct ggtctcactg gtcctcacag ttgccccgtc cccattttca 6600
gatgaggagg ctgaggtgca gagcgggatg tgggctgctc aggaccacac agctggtggg 6660
cacagagcca ggagtacagc gctgcggcag cttggggctg ggagcgagac tctctgttac 6720
cagcctctcc ctgctctgtg cctcagagga tggctgtggc cctgagcttg ggagggtcat 6780
ttgggaatct gaggaaaagt ttccacagcc tcacacagca gcaggttcaa ggtccagcag 6840
tgccaggggt ttgcatccca cccctaacgc cgctcatcca gggcccatcc gggtccgggg 6900
ctcgggatcc tgcagcgctt tcagctctgt gtctgtggcc ggtggttgtc cttcattaca 6960
gggatggtgg tgtttttcca tgttgctaag cagcatgtga gggcctctgt tgaataaaat 7020
aaagcccttg ctccggcgtg acacgggtca cgggatctct gctcactcta gtctctttctg 7080
atgccccccac cgcactgttt gccttagttt gctttttact ctctgtgttg atttctgggg 7140
catcagtcct ctgacttcgg tgttgtggct tgggcctggg gtgggggctg tggaacggca 7200
gatctgggta tgaatcccag ctccacccat ttggctccct ctcttctctc ccctgccctc 7260
ttctcccctc cccagctccc ttcccctctc atatttactt aactttactc agtaccagcg 7320
tattagtttt ctagggcttc cgtaagaaac tactggaaac ttggtggctt aaaacaacag 7380
aggtttgttc tctgatggtt ctggaggcag aagtctgaag tcaaggtgtg ggcagggtgg 7440
gtccctcctg gaggctctga gggaggctca gttcctcgcc tctgtccagc ttctggcagt 7500
ggccagtggt cctcctgctc cttggccact gtcacccgtc tccgcatggc tctgttacac 7560
ttctcttcct ataggccatc agctcttgga gccagggctt ccctactcca ggctgatctc 7620
atttcaagat ccttaaatac atcttcaaag acctgttttt ccaaataagg gcacattcac 7680
agacattggt gggcacatga tgtggggccc aattcagccc cctgcagcca ggctccacat 7740
gggtgggctc tgggtgttgc tatcatcatc cccgtttctt ggatgaggga aggccagggt 7800
tgtgtggctc atgcccagtg aggcagggac tccaggcctc ggcatgtgca ggtctgagat 7860
tgtaagtgcc atgccatctg tagagtcacg tgggtgagtg tcctcccatg gtttcctccc 7920
tggaagcggt gctggcccat ggggactgta ctgagtgtca gcgagatcct gtatggagtg 7980
tctggcagct cccacctgcc tcccctctgct tttctgtctc cagcagacac taagtgccag 8040
agcgggctct gccggtgctg gatgtgcctg accttgactt ttcttccagg caggatggac 8100
tccaacactg ctccgctggg cccctcctgc ccacagcccc cgccagcacc gcagccccag 8160
gcgcgttccc gactcaatgc cacagcctcg ttggagcagg agaggagcga aaggccccga 8220
gcacccggac cccaggctgg ccctggccct ggtgttagag acgcagcggc ccccgctgaa 8280
ccccaggccc agcataccag gagccgggaa agagcagacg gcaccggtaa gggagcaggc 8340
tgggaagccc aggctgggga tgttcaggga tagctgggtg ggaacggggt tcagccaccc 8400
ctggagggtc cccccgccag gtcctctgca gttcagcatt gtgcagctcc catgctgtgc 8460
acaggcgtcc atccagtggg gctacccacc tcctcagagc cttgcacctg tcaccttttgt 8520
ggcacccact cgaggtggtg ctggtgcccc cactcctctg cagtccttct cttcccaggg 8580
cctctgcagc acttcacagc ttccatttgc aacagcgtcc aaacatgtgg agtaaattca 8640
cgggccctct gtaaccaact agggtggcag gccagaggtg acagccacac tctcgggagt 8700
gaggccacat gcgggcgaag cccttggtgc tgagtttccc ctctccctga gggagacttg 8760
tgagggacgg gcgtccttca cctccatcct gaggcgggat gctggggccc tgggtgttca 8820
caggcagagc tgactgaggc ccttggtctt cacaggccga gctgactgag gccctgggtg 8880
ttcaaagcag agctgaggcc cttggtgttc acaggcagag ctgacaggct tgttggttcc 8940
tgggtggtga tgtgagggtg gtgggtgtga gccaggggag tccggcccct gggcaccttc 9000
tccagctgac tcaggcctcg gtcagccttc acttttgcta gagaggccgc tgccagggag 9060

gctcagtccc agctcatcac tgcccctcct gcccatgcag aagccccaga cccatggtga 9120
gtgcctgctg tgtacaccag ccttttccat gtgcctgcag gactcctcat ggcagtcctg 9180
agaggggaaa ctgagacacc gagtctaggt gacttgccca agagacccag cgggtgagag 9240
acgaggtgct cgagtggtcc aggtttcctt cttgtcccag tcaccccact ggcagggccg 9300
tggggcctcc ctgctgtggt cgcagcgcgc tgaggctccc atgtaccact tatgctgggc 9360
atgcccttgg aaactcactg cagccctggg ccctgtttgt ctttgcatct ccttaaaaga 9420
tcactgtgct cctctgcttt gttttggagg cctctgggcg gatgagggcc tgaccccggg 9480
atttgggccc agcctcacgg cttcagggga acagatacca gggctcggaa gcttggtttc 9540
aatctgggta atgagaacag cccgcaaagg acagagtcca agactgtgtg tttcagataa 9600
gaagagtcag gaagcagctg tggtttgttt ggcttcggtc tccaatgtgg gggtggtgaa 9660
gagcaagccc tgggcgccgc ctgacagccc agtccccacc actgggcctg cagctggggg 9720
cagtgggcat cccgtctccc agtcactggt cactgatgtg gccaaagcag gcctggccct 9780
gcatgcccag tggccgcagg gagggcaggg cccagggtcc ctgccagtag cagctgctgc 9840
gagaatctgt tatttagctc ccccatccag ccaacatagg tgggtgagca tcagttgaaa 9900
taccaatagg taacaacaag ggtagctatt taactgacac ttccccgagc taggcatctt 9960
tttagggcat tagcccactt gcctctttaa cagccctggg aggcaggctc tgctgctgcc 10020
ctgtttttca gatgggaaac cgagggatag agggcttaag tagatgaccc agcgtcccgt 10080
ggttggtgaa gggtggaggg gagatttaaa cataccgtgg agcctgtgct cactatgggc 10140
ctcaggtcat ctgcgcccag ggatccccat gtcccaggct cagagaagtg cacagagcag 10200
gttggcctgt ccctagatca gtgcctggcc aggccctggt gacaggagtg aaggctcagt 10260
cctgccctcc gggagccctg agcaatatgg gagacaggag tgaacaagga actacattct 10320
ggggctcatg tagcgcgggg gacaacccac attgcagtcc cctaggggcc ccccaggagg 10380
gcctggtcac tcttcagggc atcaggagga ctgcctggag gaggcatcgt cggagcccca 10440
tctgagctgc tccttggcct gtacacttgg gaaacaggag aagagcagcc ttggggttgc 10500
ccttctctgg ttgcctccct gaggaactga ccccctttcc agcccttcat tcagcaaggc 10560
tggggagggg ctgcttcaga aacgtcactg ggctggagcc ccacacttcc ttcccatcct 10620
ggctctgtgg ccttggcaag tcacctgcca cctggggaac cactgccgac gggtcggggc 10680
ttggaaactc aggggctcct gggcccagac aggtaaagga aacagtgact ccggctagaa 10740
aggagactgc agggattgtg aggctcttgg gcgtcctgtg ccagggcggg actcctcatg 10800
gcagccctga gaggggaaac tgtgtcctgt gtggagcgtg gatgcccacg tgctgagagg 10860
tgttgtctgt cctgattttt gacagcagcc caagtgatgt gccagtgggc cagggccggc 10920
acctgcctgt cggtggctgc cccagggctg atgggcaggg gccgtttctt cacaggcagc 10980
tgtggtgagt cctgtgatca cctcagtgtt gtctccagcc ctggggggat ctgttgttct 11040
gctcaacaag agaataaagc tgggctcgtc acacctggtc ccaggtgacc cagcaggcac 11100
agagaggtag agttgacacc aagcctcact cagccaccgt gacacggctg tgcgagtcca 11160
cgccctggca gagagcctgt tctgggcctg ttctgtttag ccctggggct aaagggcggg 11220
ctcagatgca cagagggaca ggcctggtgc aggtgggagg aaggaaagca gactgtgtca 11280
tcgtctcagc ttcaggctgg gcgtgctgtg gctatcatgt gcttggcctg tgctaagtgt 11340
gtgctaagca tgtggccctg gtgggcactc aggcctcctg ggcaccagcc gaggccccac 11400
cccttcccac tatgcggccc tgagctctcc tgctgtctct gtcattcacg ggtctaccag 11460
taggacaagg cctggggttg gtggaggagg gaaggaggtg agatttggga gctttccgag 11520
tcgtacccgc cactccgtct ccgagcagca gcactctcat ctctacaagt ctctacagca 11580
gtgctgccac cgccttgacc agggaatcac cctactgtga ggtggcccag ctcactgtga 11640
gactctgggg cagtcatatg ccctttctgt tgaccccagt gggtggccag cccacacgac 11700
tgctgaatct gccaggcctc ccagccattg ggggtagggg gattgctttg gcagccacag 11760
ggacacagat gcggacagga gcaggagccc tggggaggag gtgcacccca ccaggaagca 11820
cccaccaatg ctgactgagg ctgcactgcc agttagagcc gctgggcgag ggacaaagaa 11880
gggccttcgg gtgtggggaa agcaggcacc ccagctccta cactaagggc accctgggca 11940
gagagggagc ccagtggggg gcacccaaga cccatgagga attggctgtg gtgtcaagag 12000
gtcagccagg ccaagcaggg ccctgagggt cagggaaagg agagggcatt gtcgtgagtg 12060
gggctcagct gttagatgga ttgtttacat taaaggccag ccaggaggcg gagcagagag 12120
aggagtagcc aggccacaag aaggctgagg gacaggtaga catggagggg agaaagccag 12180
ggtggcctcc aggtttctgg ctggggctgc caggtggacg gtggtggcac agcctgaggt 12240
agggtcctgg gggtggttcc cggcggggtc cgagcttagt cttagagggt acaaggggac 12300
ccaggaggga ctggggctct gatgcagatg tgggagtgcg ctgagcccct gctgccctgg 12360
ggacaggcat agcaggaggt gctactcagg gcacccaccg ttgattattt aggaagcagc 12420
tgtcacgcct gttctttatc ctgacctggc ccatctgatt gctgatgtgc tcctgctgag 12480
ccccccacac cacactgagg actggccaca ggatctggtg gcactcagtc ttctgggaac 12540
ttggcagtgt gcctgtccta ctttgcgctc caagcctcgg acgtcacctc caggaaccct 12600

cccactgtgg gtgatgtcta cagtcaccta cacagcccca actctcatgg gtggacgtgc 12660
tcgttcttgc ctggcccagg gcaggccgtg gtgagctcag gggtcctgcc tgtcagtttg 12720
gtgccctcgg ctacgcaggg cctgttagaa gggtgccctc cctgccaggg cctccagcca 12780
ctgtctgtcc tgttcccgcc tcagagttca gccagcatag aaagagaggg ctgaggacca 12840
ttcagcctgt ccattcacga gaatcattgt cagtgtcaca gaggctgatg ggatgcctgt 12900
ggtgcagggt gaccccagct tccccatgtg ggggatgcgc cggtggtgca aggccaccac 12960
cctgtgccct gtctctggca cagggatgat gacagcacac agcacatcct ctttttcttc 13020
taccttccc tttggcacag ccccttgttg tgggcgtcac gagaggcgca gagagatcgt 13080
ggaggtcaag ggcttgccca ggaccagtga gggagggagg gagagggcag cagcaccctt 13140
ggagacggga ggattacatc atatttattt accacatgcc taagatcgca ctcttaagag 13200
catttgacat acatccaaaa atttttaaac actatggaag tatgtctgga tgcatcagta 13260
atatgaaagt ggaagccagc tgccctgctc ccagcccact ttgggaaaat tttgtgcatg 13320
ttttgagcca gctgtttggg aagggaggcc ccagccactg ccccttgggt gggtttgatg 13380
ttcattgtgt gggaagaagc cttccctcca tccttttcta tttctgcaga atatctgctc 13440
aatcagggag ttgagcttgg cagttcagct gctcagccag ggaggtgctg tgtcctggac 13500
tatgctagga tttcagaagg agagagggca gctgcaggcc ctatttgcag tggcttcctc 13560
ctggaatccc cgtctgctcc ctgtgtactg tccagcgtcg tagtggaagc tggtggcgag 13620
acgggtgaga cccacggaca tagatggcat ccacgaaggc tggcggagcc tgtgatagac 13680
gatggacgat gacccaggtt ccagcaagat atggcagtgg gctgtgccgg gaaggggcgc 13740
atcctggagg atgccctagg tagtgcctga gaaagtgagg gtcattgact gattccacgt 13800
ggcatccatg aatctctcaa atccataggg cccgtgcatg acaaataatt aaaaagatgt 13860
agggtgttct ccatggaacc tcacagaggg aaggagacct gcccatggta tccagctctg 13920
atcacactgc tattttgtga ttgataggca gactttcaaa gaatgagtta gattcaggac 13980
ccttttagtt gtaagtgaca aaatctcaat ataaattgac acaagccaaa aaagtagagg 14040
ggcaagggat tgatgagtcc atagatctgt agagcctaga gtagtttgct tcaggcacgg 14100
ctggatccag gggctcagat gatgtcatca ggaagtagtc ccccgccccc attccactct 14160
gctttcctgt gctgcctgtg tctcaggggc tctgcccaca gggtgacctt ccacagttca 14220
agctggcatc cacgtgccaa gatgagcaga aggtgaaagc ccatttccca gcagttctag 14280
cacaaactcc gctgggctgt gtccatccct gaaccaggtg ctgtagtcag ggagatgacc 14340
tggcaactct gattggccag gtgggggtgc gggccttctg tgcagcgcag gggagggggc 14400
agtgccaccc agaccatgag gactgagagt ggcggggttc cccaaaaggg acatcaaggt 14460
gcagtcatca gaaaaagcaa ggaacggagg tcgagcacac agaaaacacg tgtgtttgat 14520
ggttcttggt ggtccacact gatgcagccg gcttctgggg gtgatctttc tggcactgag 14580
cagtgagcat gctgggattt gcctttgagc ccacgaaggc tcctgcttca tggactgagc 14640
taactggtct gcaggacgag cccaccactg ccatatggaa gctggaataa tgaagagggt 14700
gtgtgtacat ctcaaagaga ttctgggctg ggcgcagtgg ctcagacctg taatcccagc 14760
actttgggag gctgaggcag cagatcacct gaggtcaggg gttcaagacc agcctggcca 14820
acatggtgaa accctgtctc tactaaaaat acaaaaaaac ttagctgggc atggtggcgg 14880
ccgcctgtat tcccagctac tcaggaggct gaggcagaag aatcgcttga acccgagagg 14940
cagaggttgt ggcgagccga gattttgtca ttgcattcca gcctgggcga tgagagcgaa 15000
actctgtctc aaaaaataat aataaaaaag agagagagat tctgtatatc ggggagatgt 15060
cttgaaatgt caggctgtga ataacttctg tgaaaaatta attatggact ggaagaatct 15120
ttttagacca ccatatagac ttggaccttc catgttttga ttttttaaa aaagggaaga 15180
ataatatgac ttgtgaactt aggtcaagta aggagacttg aattcactag aggtgctggc 15240
gttgtggtgg gaagtacagc aacagccctg gggtctcctg gggaagtggg ggagggaaga 15300
gaaagcctcc tcttgtgaac agcaactttg gaccactcac tacaaagcat ctcttttcat 15360
cataaaccat cctatcagat tgatgttaac gatcagattg aagctgcttc ggagcagtga 15420
tgtggcctgc accaggttgc acagcacctt actgtggagc atggatccac acgtgcgtct 15480
gtgtcattgt gatcgctcac gctgctctgt gggcccatgg ccttggtgcc tgcaggctct 15540
gtgtctgcag ttcctctagg gatagtgagt ctgctgctga gcccgcagcc agaaaccacg 15600
ctgggggggct ggcagggtga cggcagcaag gaaggagaac tgtccttgat ggattcaggg 15660
agcaccaagg tctgggggct ctggggaggg tgggtgtcca aaagggctgc cctgcgggac 15720
ctgatgccat ctctgagcag cgtccttgag gaccccggca aggaagctga gcccgaaact 15780
ctgtgcaaga gttgcgtcca gggagagtta catccaggga gagttacgtc cagggagagg 15840
ggcgtccagg aagaggggca ttcagggaga gggacattca gggaggagtc ttccagggac 15900
ggggacatcc agggagaggg gcgtccagcg agaggagcct ctagggaggg gtacatccag 15960
ggaggtggcc tgtcagggag ggtggtgtcc agggaaagca gcatcaaggg agaggggtgt 16020
ccagggatgg gggcatccag ggaggggggc ttccagggag aggggccttt gtgttggcga 16080
gttgtccatc cacagggctg gagtgcgtgt gcctgaccac gagagcttgt gaagaatcca 16140

tctgattctt tccttctcca aacacatgcc tcacccacac agccttacct gtgagggcct 16200
cacctgccac ctgcgacttt tgagttgtcc cagcctccct ctgccctctc caaactgggt 16260
gtgacccttc ctctctgcct ttgcacctgg ccacatccca tggtctccca ctggagtcct 16320
gtctcctgta cacctgggtc atctgaaggc agtggccagg cagcacttgg tgctgcagag 16380
ccggttcaga gcctggcacg gtatggttct gggaaagccc tgggtgagtt ggaccagaaa 16440
ggctggtcct cagtgtggtg ttactgggaa ccatgcagcc tgtcccgttg tcagccagcc 16500
ttgggtgaaa tgggtaagac tgcatgatgc ttttgtaaat tcaggtaact tcttaggaag 16560
taaaatgcag cacaaactgt agttgatgtt ttgacccttg gtctaataat tcaactggag 16620
aacctgtgct aggaaaatga cttcccaaaa gaaaagaaac ggttcatggc accaccgtag 16680
aaaacaagag gttattggaa gcaacctaca gacctgatgg agaggagggg agggtcagta 16740
acctgactct gtctccttga tggatctagg agccaccaaa atggtgagct gactgcagcg 16800
catcgtggga agccagagaa caatccactg ctgtggactg aaagtttgtg tctctcccag 16860
attcatttgc tgaggtcttc acccccagtc ccatggtgtt aggaggtggg gcctttggag 16920
gtgatgaatc atgagcgatg gctttacggt ggaattagtg cccctataaa agaaacctca 16980
gggagctccc ttgccccttc caccctgtga ggacacagcc agaagacctg aaacaaaaag 17040
caggccctca gtagacacgg agtctgcaac actatgatct tggacttccc agtctccaga 17100
atcgaaagaa atacatttct gatgtttgta agccaccagc ttggtctatg gtattaatat 17160
ttttgttata acagctcgaa cagactaagg cattcactct taagaaggaa aaggtaatat 17220
caagatgtat atacactgat tataatatcc aaaaaatgtg gtttttgttg gtacaatcat 17280
ggaaaacatg aacttgattt gttggggttc tgggcttcgg gtacaggtgc tttgtattat 17340
ttaggtcatg ggctgtcaaa cttgagtcct tggatctttc tagagcccca gctagagcag 17400
gcctaggcca aggggtgtga tggtgggact tggctctgtg ctcgcctcct tcccatgcct 17460
cccaccccga gcgctttgag ttgctgttga accacctcac ccgcatagag tttgctcatc 17520
ttgctctggc gctgcttgtg accggaaggc aaacagtggc tttctgggca tcttccctgt 17580
ctgtgtctgt gtctctctgt actcctggcc tggatttgaa agtgacaagc agcagcagat 17640
ctgaagaccc tcatgccttg tcccctccat cctgacaagt gacaaagtct ggctttgtga 17700
catgtgtgtt tgtttcttct gtgttaaagg gcctacaaag ggagacatgg aaatcccctt 17760
tgaagaagtc ctggagaggg ccaaggccgg ggaccccaag gcacagactg aggtgaggac 17820
tgcggtgccg gcagggactt cgggacgcgg ccccggcaca acaggcctgg ccacgagctc 17880
cacagcccac agagaagtgt cggtgcctga gatcggggtc aggagccagc gtggtgcacc 17940
ctaccccact tgagccccat gttggtaggg tgcccatgtt cactgtgcca gttttcctcc 18000
tggcactcct ctggggagca gcgctcatcc ccctttgtc caactcacac ctcatcttgg 18060
gcatcacctc ctccaggatg acctcctggc ttcctgcagc tgcctgctca gtgccacccc 18120
tcaacataca ctgtgtatgc tgcccactct tgtctccctg actggccttt cacctggctg 18180
ttagctgtgt gcacggggcc ctcagagcgg tgaccattca cttgggcatc agccaagggc 18240
tgggctttgt gccagtgctg gggacatgat gtggccctct cttcatgggt gacaggttag 18300
tggaagaagc agacccaaaa aacccaagta gacaagacac agaaatacag tgattttgaa 18360
ttctggtgag gggagaagcc agcagaacgc tgagacggga ggtggcaggg ggtcctgtgt 18420
cctctgtgga ggcaagttct caccacctca ctctgtcccc ttgcacaggg ctcaagcaag 18480
tgaccacagt tatcctgagc agtttcctgc cccctcttct ccctaaagag gagaaaagct 18540
tccacgcgaa accataaatt aagaatttct ggctgggagc ggtggctcac gcctgtaatc 18600
ccagcacttt gggaggctga ggcgggtgga tcacctgagg tcaggagttc gtgaccagcc 18660
tggccaacat gatgaaaccg tgtctccact acaaatagaa aaattagcca gacatggtgg 18720
tgcatgcctg taatcccagc tacttgggag gctgagacag gaaaatcact tgaacccagg 18780
aagtggaggt tgcagtgagc cgagatcgca ccaccacgct cccaactgag caacagagca 18840
agactctgtc tcaaaaaaaa aaaaaaaaaa atttctttcc tgacctcaca gccatattgt 18900
actttaaagt tcctcccaca tccttttgcc caggcatttg gaggcggggg gagcatagta 18960
agccctttta tgggctctct gtgccctggg cctcacttt tccatctgta gaatgagaag 19020
actggaccat ttcagcaggg attctttttg caaagggcca gatagtgaat attctggact 19080
ctgcgcaatc tctgccgtgt tgctcagctt tgctgtggag cacgtaaaag cagcatggac 19140
aacatgtaaa tgaatggtta tggttgtatt tcaataaaac tttatttgca aaaacaggca 19200
gtgggccagt gctaaccctt ggttagctgg cgtctctcca gccctggctt tctatggtcc 19260
cctgactgtt ttgagaggtg taaaggaccc agtcatggtt ctgatttcca tgcattgatg 19320
gtgagccttg gcaggcagga gcaactcaag gaagagaacc tgtaccagta ccagtcggag 19380
cccgtgtctc cctcgccgtg tggatggggt ggccacacct tcctcaccgt gttttgagga 19440
gcgagtggcc ggaggctcag tagggcctag cctagtggac atgcctggtg tgaccccatt 19500
tctgcccctt ccttcctggc ctgggtgaca aagggaagtg ggtgaaagga ggtgggctgg 19560
cagggagcat ggggtgggag agggtcggag aatctggagg ctgactggtg tctggcttgc 19620
aggtggggaa gcactacctg cagttggccg gcgacacgga tgaagaactc aacagctgca 19680

```
ccgctgtgga ctggctggtc ctcgccgcga agcagggccg tcgcgaggct gtgaagctgc   19740
ttcgccggtg cttggcggac agaagaggtg ggtctgtgtg aggcttagaa cagcctctgg   19800
agggttgagc agcttgtaat gctgcttgct aactgaacaa ctaaaatctt accaaaccta   19860
acgctggtga tgctgttggg aaatttcagt ttctgttttg ctggtggcct tctcatttta   19920
gacactgttt ctggacttaa catgggatat ttaacagacc aagccatttt cattctcttt   19980
ggcttgtgtt ggtatctcca agtggtaact atacatcctg cttccctgct gggttctgat   20040
ccaaactgag acatcgatcc tgggttcagc tgagaggcag tttccctgag gcaaagctgg   20100
agactgtaat gtggcctgca gtcagagtgg cccatgcccc tgagctgatg tccgtcgagt   20160
ctgggcagac cagctgaggg gctgtgggcc cacacagacc tggacaagca tgttcacttt   20220
gggaggaggc aggctgaccc ttcatagata atctctttca aacagagagg caggggagga   20280
aagggacgca gctcaggcca ggggtggtga gttaggacgt tgtcctgcgc aagggaagtg   20340
gctgtccctg gtgccgtggg gcagtgtgct tcctgaaggt tccacgcctc agggaaagat   20400
gaaagtgggt gtcacaaagg tgtttttttct aatgatacct ttttttcag attgtaagag   20460
tattttatgc ttatcataga aaacttggac tatgtagaaa acataaagat acaaattact   20520
cataaccttg aagcccagaa gcaactcccc atattttggt gtatttcctt ctgtgtttgt   20580
tgttgttgtt ttgagatgga gtctcactct gtcgcccagg ctagagtgca ggggtgcgat   20640
ctcaagtcac tgtaacctcc acctccctgg ttcaagcaat tctcctgcct cagccttttg   20700
agtagctgag attacaggcg tgcattacca cacctagcta atttttgtat ttttagcaga   20760
gacagggttt catcaggttg gccaggctgg tcacaaactc ctgacctcag gggatccacc   20820
ctccttggcc tcccaaagtg atgagagtac aggcgtgagc caccatgccc agcttgtatc   20880
atattttaag agttacattt attatacaga gtcttctctt tgaaaattaa actgatgatt   20940
ttgaaaaact aacagtggca ttgttagcca gaatttctca ccaacccagc ctgctgattc   21000
cagagcagtt ggaatgacct ggaggggct caggcggccc acacggctca ttgttgggtg   21060
ccagggatag cccagaggct gctggggcag agcgtgcaag ggaggaggcc ggttggcccc   21120
gtgttccccc agctgcacac accttctgat gtgcagttgt gctttctgat ttgaaaagaa   21180
atacatgctc atgtggaaaa ttttggaaat gttcaaaagg agaaaaacat ccagaccccc   21240
gtgccagtcc tcctgccctc tgcttggttg tgccagtgtc ccaggccgtt gggggtaaag   21300
ggaggaaggc agctgctgac cagcggtaaa acctgctctc cctctccccg ctgtccaagc   21360
cttccttaga aagggctggg ctgtgtgtga gacagaggaa agctctgtag acacaggggt   21420
ggccctgcca gtgcacccag gctggtgact gcagtgagtc caccctgggt tcaagacagc   21480
tctgcctctt tctgactgtg tgaccttgag tgagccagaa cacaggcctg tgcctccggg   21540
tctgcacctc tgagagaggg gaggaacagg aagaggctgc cttcttcccg ccagcatgta   21600
gggggcactt gggggcgtct ggtgggagac cagtctggcc tcccagctgg agagtgggcg   21660
tggcgcgatg tcctcttgag tcagatgtcc atgcatcctt ccctggtaac caagtcctga   21720
caccttctat gagtctcgct cgaaagcctt ccaggcagag ttggcagggt cagagtggca   21780
ccgaaagcct aggcagggca cacaaggcct ttgaccacat cctatccctc aggcatcacg   21840
tccgagaacg aacgggaggt gaggcagctc tcctccgaga ccgacctgga gagggccgtg   21900
cgcaaggcag ccctggtcat gtactggaag ctcaacccca agaagaagaa gcaggtggcc   21960
gtggcggagc tgctggagaa tgtcggccag gtcaacgagc acggtgcgag gattcaccct   22020
gggcaccagc cttccctggg cgccagcctt cccacaggag ccaggacctt cccatagggg   22080
ctgggacctt ccctcagggg ctgggtcttc ccacaggagc cgggaccttc cctgtgagga   22140
cagggccctt ccttgtgggg accaggggac cagaaccttc ctgtagagac cgtgccctag   22200
tggtgaggtg tgtgggtggc attttgacag catctgccct ggctcaagtg ctcactcatt   22260
gaataaacca gagggtattc tgcccagtgc tctgtgacca cgtctaccaa tgggacggac   22320
tgtgtccatc accaagtggg agcacgctac gtggtgctga gtccacccca gctactggag   22380
gtacagaggt gtggcccctg ctctgcctgc cctgggggcc ctatgatccc cagaacgtag   22440
gatgcccctg gaactggcgt gccctaggaa cagtgcgcca gtttctggtg ggctgcaggg   22500
cacgaggaga tagtcaactt gtctgactgt taatccaccc tgtcccctgc agatggaggg   22560
gcgcagccag gccccgtgcc caagtccctg cagaagcaga ggcggatgct ggagcgcctg   22620
gtcagcagcg agtgtgagtg cagcccctgc cccgtctcac ccatgcctcc cagcctgcac   22680
ctgcagggcg acctctcctt cctgtgcgac tccatcctgg cctgcctat ctcacccgtg   22740
cctcccggcc tgcgcctgca gggcgacctc tccttcctgt gcgaccccat cctggccctg   22800
ctaggatctc aggcggtccg tttggggctc agtgttctgg acgctgggag tagaccctgc   22860
ccacctggag cgcacgcact ggagggaagg cagacccagg acagaaacca tgatgtgcca   22920
gtccctcttg gacaaggaaa tactgggggt ggggactggc gggggggtcc tgagtgggga   22980
aggatgagga gggcattagg ggaaggggct ccgggcagag ggaacagcat gagcaaaggc   23040
ccaggagacc cagcagggtg tggaaggggc tctgccccag cgtaggccct ttgtaggtag   23100
gcagggaaca gagtgggcag agcctaggtt gggggataga gtctctgggg cacactgagc   23160
tgggggtgcc tctgggatgt ttgtggtcaa agactgagaa gagacgtcac atcctccacg   23220
```

tgcagctggc cagcctcact gtcttccatc cacgtggggt agactctgca gtctgaggca 23280
ctgggagctc ctgtcctgga gaagatagac aggcaggagg cttgggggct cagtatcaat 23340
ggggggatgg gccaagaccc aggtctgcca cagactcttc ttgtgacatg ggcaggtctc 23400
atttttcttt ctgggtgtca gtttcccctt ctgaacagtg agatgatcag acaaggggac 23460
tgaaggacct cgcccatcct agaagcctct gtggcatgta gggcagagta caggtgggag 23520
cacagatccc accttcccac tgggcccgtg aggaaaccgt ccatagggtg tggtcagagc 23580
cctcgcctct gcgtgggcgg acatgggtgc ttgcgctgct gttttcctgc tccatttggg 23640
ggagacaaag aaacacctca catggaagtg ggaagaaatc tgtgactcta ggaagcccga 23700
ggaccagaga ggctgtgaca gtgtcctgcc accgaggtgg catggtgcct gctgtcctga 23760
agtgaggtgc ctctgggcag gagcagccca gcctgggtct ccatcctggg gggcactgca 23820
ggcagagtct ggaccccaac ctcacttccc agtcccacag gggtttgtga tgagccgggc 23880
acacctgctg ctgggcctgt cgctgttgct gggggcgagt taggagtcag cgctgccagc 23940
ctgtgtcccc tccgccaggc cctcagggac ggggccccag caccccatt gttacccctc 24000
tatcagaacg ttcccatcag gcatcatctg tttctctgcc agcttttcca acagaactcc 24060
cagaaagggg cgtctgtcat ccccttctct cgtcctccgt cctttgctct caggcctctg 24120
ccacgttatt ccaccaaagc ggctcgtcta ggccatcagt gacccccttg ctcagtcact 24180
tcttgtcctt gtctgacttg atctgtcagc agcgtgggca gagatgactg ctcttctcct 24240
tagagccctc tcccccacca ggcccggggc ccttgctgcc acactggctg ctccttcttg 24300
gggtccctag actgatgtcc tgtctctcag gacaggacat cctgcccaag gtattcatcc 24360
ctatgaggac aggggccttc cttgtgggga ccagggtctt cccgtaggga ctgtgctctg 24420
gtagggagat gtgtgggtgg cattggtgac cactctcgtg tgtcagacgc cgtcctgaac 24480
ccgcaagccc ccagcccgcc caggctgcat ccagcccgtg ctccttccac atctcctgta 24540
cctggctgtc ccatggtcac ctcacagctc agtctacagt gagctactga ggcctgtgct 24600
acctgccact tgtgtgggcc cccccaaggc tgcctcatct ctttgctgca gctacgcctg 24660
ggtgctgccc cgactcctct cctcttgcac acacagctga tccattagga ggtctgtcag 24720
cgcccactat cacggctgcc atcttggatg cagcctgcct cctccctagg gtttctgcag 24780
caccctggct tgtctccctc tgccctgccc cagcatggtc ctttctttcc catccttccc 24840
ctggaggccc tcctggtcac catctcacca cggcaactgg ggtccttcag aggcctcatt 24900
ccctcctagg ggatcatgtt ggatagaggg gtcagtgggc acgtagctct cgcggatcac 24960
gttggatgga gggatcagtg ggcacgtagc aggctcttgg tggaagatgt tggatggagg 25020
ggtaagtcgg cccatggcag gacgcatgtt gaatggacgg gttagcaggc atgtagtagg 25080
cccatggcag acacttgttg gatggaggtg tcgcaggcac atagtaggtg tgtggcagat 25140
catgttggat ggagcggttt agcaggtgca cagtaggccc tgggtggatc acgttgaatg 25200
gaggggttgg gggcgcagat catgttcgat ggagcggttg gcagggtgct gccctacagg 25260
gcggcttgtc acccgtgctg tgagaagccc cacgctggct gaagaggagg gtgtcagcat 25320
gcactggagg tgcatgttgt aaggcttggc aggtgaactg ggtgaaaggt gtgggtgagt 25380
ggccccaggc ataaggagct aggcagagag ggacacttgg ggtgggggac agcacaccca 25440
ggggccgggg gccaggagtg gaggctggca cttggcaaac ctgcccccct cctcctcacc 25500
cagcctggtc ctcaaccctc aggccgccca gggaagggtt tcctccacct gaacccactc 25560
agctcctttc ttagcttggc cccacgccac cgtccccagc ccattgctct gtgtgagggt 25620
ggcagtgggg ctgcagtgtg gggcgcccat gctgttttct ctcatgcttc agccaagaac 25680
tacatcgcgc tggatgactt tgtggagatc actaagaagt acgccaaggg cgtcatcccc 25740
agcagcctgt tcctgcagga cgacgaagat gatgacgagc tggcggggaa gagccctgag 25800
gacctgccac tgcgtctgaa ggtgagtgac caagaccccg gtcaggccgg agcctgcctc 25860
ccaaggactc gcgcacctca ggcagggcac cttccaggaa gctgcaggtg gggaggttcg 25920
cgcctaacaa agagtgtctt acagccgtgc cgctggtacc tttgggtcat catctatcgt 25980
cataaggatg tgtcctcggg agagaggcct ttctttttctg cgccgtcagg ctcagaaacc 26040
aggggcggtg ttgggcagga gtgctaggat ggcaagcaag gggcccctgg gtctttctgt 26100
gcagtgtagg gggcagtggg gcgggctgga cgcagaattt catctagact tcatcactta 26160
gcagatttct ccatgtatgg tccagagtgg aggccagtgc cttgaagatg ccttccccga 26220
aaacaaagtg ccgtgtccgt caggccaggg aggacgtgca gaggcgccta ctgcccctgg 26280
atctgcaagg gctcagggag cgaggccagt gtgggagcgt ggtgggtgac acttgtgcag 26340
ggtgccattg ggcacacctg gaggacatga gcagctttgg cggcaggagg aaagcgggga 26400
gcggggagcg ggaagcccag cagtgcgggc cctactcctg gcctgacagc ttgttttctt 26460
ggccatgtgg ccctatggcc tctcacagag ttgctcacag ttcagctcgc atggttgaat 26520
tagatgtgtg cctgtctcac tctgcagaga ccaccaggat ttccagaagc ttggcctcct 26580
gacccaggat ccccaagctg aggccagcac agggggcccc agcagatgat tgtcacagag 26640
gctgcgggca gtatcatgca ggcaccctgg ggttctgtct tctgaactga ctgagaacct 26700
ggaggaactg ccatcattgt attcagatgg gcgcgtgagt cagggccagg accaggagat 26760

ggttgggcca ccaactgtca ggcaggcagc aggctgaaag gtcaccgtcc tcgagccagg 26820
gagaaggtca gagtaggccc gccctccggt ggggggggaag agcaggaacc caggtgctgc 26880
aggagcagct tcagcccaaa gagcagcagg cacagcagcc ctgctgcagg gactcggagc 26940
cagacaggtc tgagtgtgag ccccccgagt gcccagagct gcgagtcctt tagggtggca 27000
ggtggaaggc tctgctcctc ccacttggga caggctgtat gtccagggta tgacaccgac 27060
cggtctgcag aggttctgag acatgggcca ggatggccaa gaacttgaac caagggagca 27120
catggttttt actgatggtc tgctcagagc tggccctagg caagcatatt cacagattgg 27180
gtgccatcat ctccttacta atgggagatg aggaagctgg ggtgctggca tggttttggc 27240
aaagacactc ccgaaatgag agtgagggtc acaggtgaac gggttaactt ggagaagacc 27300
cttgggccag gcggccatgg cccaggtgtc ccagtttatg ccatctggtg taagtactat 27360
tataataatg cctccttgca tatgcgtcct gttcggaaaa ttatgtcacc tagttatgga 27420
tggtgagtct agtgttgtgt agcatctgaa aatcctattt atttatgtta cttaatgaca 27480
gggtctcgtt ctgttgcaca ggctggagtg cagtggcacg gtcacagctc actgcagcct 27540
caacgtgcca aactcaagtg atcctcccac ctcagcctcc caaatagctg gactacaggc 27600
atgtgccacc acacccagct gatttttttg aaaacatttt gtagcgatgg gggtctcact 27660
attgttatct acgctggtct caaactcctg ggtcaagcga gcctcccact ttggcctccc 27720
acactgctgg gattcaggcg tgagacactg tatccagcct gaaagtcctt tttagaaggg 27780
agagatttaa gctgttttga caagaagtaa acaagactct gtttcccttt agttttggat 27840
gctgtcagag cacttcgatt cctttgcagt ttgtcttctt agaagtctct tgtgagggag 27900
gcaatgaatc gtcacaagcg tctcccaggg ctgagagctg tcccaccttt tagtagcctt 27960
ttgtagaaca aggagaggtc ttcactcaaa gtgaaaatga cagacacctc tttgtgagcc 28020
acccttactc ccacaaagaa tgttccagac tccacacagt tctcttctgg agcctggagg 28080
tgcaagctgt catcgtagat agaatcctct gaaaacaaac ttaatatctt aaaatcgtaa 28140
catttttaac attacataat tggaagatcc gacacgggaa catttttgcc cacaggtttg 28200
agcagaatct cgtctcggct cccttcccca gcctgccttc ttcacgggtc cgctgggcca 28260
tcggtgtggt gacgctgctt cctggtttct catggcagca gaagggccga tcaccacagg 28320
cccagacaga gagctctgca gtgttctatt catagaggca gactcctcat agtttttaag 28380
tttcgttctt tttccatctc aggagttgaa tttgtttgga ccttctttgc agaaaaggag 28440
actgaacttt ttaatcccat gtcagaggaa ggtcttgggt gtacattgca gggagcagaa 28500
attggggctt taaaaagggg atcgttccct ggtcatcgag ttcaggtggg ggaccagccc 28560
ctccacactg tccaagcaga ggctagacgc catccaacgt aatgttctag aatattcctt 28620
gtcttttaaa ggccctctca accaggtggt tgtgtgtgtt ctaggcaaaa gaatgtgcat 28680
ggctttgagt aagacaggat ttcagattta tttttaccaa ggccacccca ttgctaccgt 28740
ggcggtcgtg cgtgaaactg gtggtgagca gtggcgggcc aggccacctt tagttgtcct 28800
gctgttcgca aggcccctgt acggggatta actgagcagc ctgccggctg acgcacatgt 28860
tgactcgtgt agctgttgat tttagccagc tggcctggcg tcatactaca ggggcttgcc 28920
ccaaacaaac cctggagtag gttggttttc ataaatgccc tctggctctc tggcaagcgt 28980
tggaaatgcc ctcaaaccct gtagtgcctg gtcttccctg gccatctcct ggcagtgtga 29040
gcgcctctct gctcatgtag aaggttgttg accccagaga gagctagggg agctgttagc 29100
tgctttaatc gtttatttgc ttgttataaa acagagatta agtggactgt cacaaatata 29160
catgaagcag cagagctgtt cccccacgtg tagacgtgca tgcacacact cctaaacact 29220
tccgtttctt ccatgagcct gtccttctca tctgtggaag gctcagtctg tgtagacatg 29280
catgcacaca ctcataacac ttccgtttcc tccatgagcc tgcccctctc ctctcatctg 29340
tggaagactc agtcttctgg ttgtttccag gtgtcctggt cctccagagc tgatagccag 29400
gctgttggca ggtgagggcc acattctggt gttgtcacct agttgccatt cacctgggat 29460
ccctccttgg ccttctggtg ggcctggtgg agagaacctg ccccagggga tttgtggggt 29520
ctggtgcttc cccgaggccc ggagctgatg gggggctcca gcctggatgt gtcatgtagc 29580
ccctcagtgg gccctggaca tgcagccagg cttccctacc cccatgtggg gccgcccccc 29640
aggcgtccac ttcctggtgt tggtccaggc ccctggcact tctgaagatg cacagaacct 29700
gctttgtatg ccacgccatg ctgtgtgagc actggcacca tggccacttt gtcccccatc 29760
ccctccagtc cctctgcctg gctgccctgt gttctggcag ggctcagggc ctctgagttc 29820
tgcacagacc aggaaggcat gagggctggc cctgggtggc agcgctgctt caccccatcc 29880
tgaggctgct cccagcatgc ctccctcagc cctctgctcc tcttggctgt ggtgaaggcc 29940
aagagctctt tctgaccttа gcactgcatg cagagcaggg gagcaggacc cacaggtcgg 30000
gggaaggcag cctttggcct tgactatggg gaggtagggg tgcactggtg tgtatgtgta 30060
ggtgcacacg tgtaggggtg ggtgtgcatg tgtgagggtg cacgtgtggg ggtgggtgtg 30120
cacgtgtgtg taggggtggg ttgcgtgtgt gtgaatgtgt ataggggtgg gttgtgtgaa 30180
tgtgtgtgtg taggggtggg ttgcgtgtgt gtgaatgtgt gtgtaggggt gggttgtgtg 30240
aatgcaggta ggttgcgtgt gtgtgtgtag gggtgggttg cgtgtgtgtg aatgtgtgtg 30300

tagggtggg ttgtttgcgg gtaggttgcg tgtgtgtgaa tgtgtgtgta ggggtgggtt 30360
gtgtgaatgc aggtaggttg cctgtgtgtg tgtaggggtg ggttgccgtg tgtgtgaatg 30420
tgtgtgtagg ggtggggttg tttgaatgcg ggtaggttgc gtgtgtgtga atgtgtatag 30480
gggtgggttg tgtgaatgcg tgtgtgtagg ggtgggctgc atgtgtgtgt gtaggggtgg 30540
gttgtgtgtg tgtgtatggg gtgggtttgt gtgaatgcgt gtgtgtaggg gtgggttgcg 30600
tgtgtgtgtg tgcacgtgtg tgtacgggca gaaggtggcc tggcagctgc ttctcagggt 30660
ctccggccat agggagggca gctcacccgg cagccccgtg cagctgaacg gtatttttgc 30720
tgggagagtg gtgctgggcc ctctgcctgt gtctgcactg aactcccttt ccctaatgct 30780
tgctctgcca cccctccggg gactgcctgc ggcttctcct cctcgtaagg atcgccattt 30840
cccttcatgg aggctcaggg aggtgacatc cttgctgtag tgcagagctg ggtcacaaaa 30900
ccagctccca gacaggcatc tcagggctcc gtcatcttct catgtgggtg accctgaggg 30960
gagggaagtg gggagcccgt gggtccctcc agtgctggag gtcttgcagg gagagaagca 31020
cacatgcatc tagtcacgct ggtagaaggt ggggagccag gcacggggca gaggggggct 31080
ccaggcccag aagagggagg gctcacaggg actgcgagca tggggagggc cacgtggaga 31140
agggggggagg gaggaccact aggatggggc tggtgatggg aaaacgcaag ggtgcgggtt 31200
ccttttgccc agaggcaggg tggtcagagg gaggcgtgag atgggagcag tgggggtcct 31260
gtcccagcct cgttcccacg taccatcttt cccccaggtg gtcaagtacc ccctgcacgc 31320
catcatggag atcaaggagt acctgattga catggcctcc agggcaggca tgcactggct 31380
gtccaccatc atccccacgc accacatcaa cgcgctcatc ttcttcttca tcatcagcaa 31440
cctcaccatc gacttcttcg ccttcttcat cccgctggtc atcttctacc tgtccttcat 31500
ctccatggtg atctgcaccc tcaaggtgtt ccaggacagc aaggcctggg agaacttccg 31560
caccctcacc gacctgctgc tgcgcttcga gcccaacctg gatgtggagc aggccgaggt 31620
taacttcggc tggaaccacc tggagcccta tgcccatttc ctgctctctg tcttcttcgt 31680
catcttctcc ttccccatcg ccagcaagga ctgcatcccc tgctcggagc tggctgtcat 31740
caccggcttc tttaccgtga ccagctacct gagcctgagc acccatgcag agccctacac 31800
gcacagggcc ctggccaccg aggtcaccgc cggcctgcta tcgctgctgc cctccatgcc 31860
cttgaattgg ccctacctga aggtccttgg ccagaccttc atcaccgtgc ctgtcggcca 31920
cctggtcgtc ctcaatgtca gcgtcccgtg cctgctctat gtctacctgc tctatctctt 31980
cttccgcatg gcacagctga ggaatttcaa gggcacctac tgctaccttg tgccctacct 32040
ggtgtgcttc atgtggtgtg agctctccgt ggtcatcctg ctggagtcca ccggcctggg 32100
gctgctccgc gcctccatcg gctacttcct cttcctcttt gccctcccca tcctggtggc 32160
cggcctggcc ctggtgggcg tgctgcagtt cgcccggtgg ttcacgtctc tggagctcac 32220
caagatcgca gtcaccgtgg cggtctgtag tgtgcccctg ctgttgcgct ggtggaccaa 32280
ggccagcttc tctgtggtgg ggatggtgaa gtccctgacg cggagctcca tggtcaagct 32340
catcctggtg tggctcacgg ccatcgtgct gttctgctgg ttctatgtgt accgctcaga 32400
gggcatgaag gtctacaact ccacactgac ctggcagcag tatggtgcgc tgtgcgggcc 32460
acgcgcctgg aaggagacca acatggcgcg cacccagatc ctctgcagcc acctggaggg 32520
ccacagggtc acgtggaccg gccgcttcaa gtacgtccgc gtgactgaca tcgacaacag 32580
cgccgagtct gccatcaaca tgctcccgtt cttcatcggc gactggatgc gctgcctcta 32640
cggcgaggcc taccctgcct gcagccctgg caacacctcc acggccgagg aggagctctg 32700
tcgccttaag ctgctggcca agcacccctg ccacatcaag aagttcgacc gctacaagtt 32760
tgagattacc gtgggcatgc cattcagcag cggcgctgac ggctcgcgca gccgcgagga 32820
ggacgacgtc accaaggaca tcgtgctgcg ggccagcagc gagttcaaaa gcgtgctgct 32880
cagcctgcgc cagggcagcc tcatcgagtt cagcaccatc ctggagggcc gcctgggcag 32940
caagtggcct gtcttcgagc tcaaggccat cagctgcctc aactgcatgg cccagctctc 33000
gcccaccagg cggcacgtga agatcgagca cgactggcgc agcaccgtgc atggcgccgt 33060
gaagttcgcc ttcgacttct ttttcttccc attcctgtcg gcggcctgag gatggtccgc 33120
cacgaggagc ttccagtgca tgttgccatg aggcccttcc ccagtgtggc cccagcccga 33180
caggcatgca ccagtgccat ctgtgcccac gtgtgcagac tgtggctgca gagaccttgc 33240
gaccatgtgt agattgcatg gaccccgaca aagggaaggc tgctgtgtag ctctgtccac 33300
tctgaatacc aagtgtgttg ggaattgcat gccatctcca ccctgagcct gacctttctg 33360
aatgacatgg gtgtgccagg ctggactagg aggttccggt gtctggaaaa gcactttaca 33420
gatgagattc cctctcctcc cccaccttca agcaccctgt tccctctttc tttcttttgt 33480
gttggatttg tttaaaaacc aaataagcat ctgtgtaacc tccacagtag catttcttat 33540
ttgtttggtc actgctacac cttagcagct cttccccttt cctgggggat gtgcacggca 33600
gcttgagcct gtcacgtggt caaggcccgg ccccatcaga ggctgggggа ggcggcacat 33660
tggcagtgtg tcacactgag ctgggcacca caggctgcct catgaccctc ctctccagca 33720
ggtagtgggt gaatgtgtga aggtcttgcc tgaatccatc aggacttggg aaacagagaa 33780
ccctgtgggg gtggctgtgg gggaggtccc tgccagtgtt tagaagagcc tgactgtgtt 33840

cagtgccttg gagcagaaag ccagggtcct gagtggctga aataaaagcc tctggtggaa 33900
cctgcagcgc tttccttcct ttctttaccg aaaagaagtc tttcttgtac gtgcgtgaga 33960
atcagcagag cctgcactcc tgttgaatga aatgcaagtg caatttgagt tataaaagag 34020
caaggttgat gtttcacagt tgatggcttc ctgccacagc gagaccctgg cttcatctcc 34080
agctggaggg gcccctgggg catctgccgt aactgtgggg tggcctgggc atgggctgcc 34140
tgtgcagaga gacctgtgct gaaggtgacc atggagtgtc agcccagcca tcttcagatc 34200
ttactgagcg gaaagtgcac ggtcataatt atctttgaat tccagttacc catgaagtgg 34260
agcctacagg gtgattgatg gcgtgggggc tgctgtccat gcgggagtgt gttgtgaaac 34320
cacctcctgt gttccctaac ctgcagtgat gaccccatgg gattataatc agaacttagt 34380
catttggggg cacccgagca gctgcaggcc cctcagcaaa cctcaccccc cttgggcagc 34440
ctgtctgtcc cagccccttg ggtctttggt gtcttccctg aggttgagga cgtggcctgt 34500
acctgctcct gtccccccgta ggctgggctg ctctctgctg ggcaccaggc ctcacctctt 34560
tgctgtgggt gggaagcgtc gagtggtcac acgatctggc ttggccacct ggtgttgcca 34620
tgcaggatgt ggttgctcca ttttccttgg acgacactgg gtggtggcag ctacacccga 34680
cactgtcctc tcatcctggt ttgtcctgcc tggaggggct gcgtgaaacc acacgagagg 34740
ctcaggtggt gtcatcactc cgagccacgg gtggtggaag ctgcaccccc acactgtcct 34800
ctgcgtgaaa ccacacgaga ggctcaggag gtgtcatcac tccgagccac atgctcgccc 34860
ctgggtttct ttcccaggca atgagatgag gttggggttc agcctcattt cctctcccca 34920
ttctgtggat ttgtgaagcc atggaaggcc ctcgccctgt gactctgccc ccaggatccc 34980
cactgactgc agaggaggga agtcgcctga tgctttccaa gtgaacaaag atgacaacac 35040
agcatctccc tcacaacacc tggatgagga agatgctcca acatccccca tctacagaag 35100
acggcaccga ggttcagagt ggaggccata gcacattagc acgttatcac atggcacata 35160
gcagagcagg aaggacagag ctctgcctgt ccccttcccc tgggccacca cactgccact 35220
cagcatccac aattagtttc taaccaaccg gcagaggcac aggttggagg agaagtggtg 35280
ctgtcttttg ggcagagtcc agaatgggca tggtcccggg ttccactcat tctgccacaa 35340
acccctgtgc cacccggcac catttaacac ctgtccccag gcctcagtgt cccaatctgt 35400
aaaacggaaa gcctgggcta agttagtggt tttcaaaact tcttagcaac accgttttgg 35460
tttgtttgtt tgtttgaaat agggtctcac tttgtcaccc aggctggagt gtagtgctat 35520
ctcggctcac tgcagcctcc acatttgttc tgggttttac agagaggctc cccgggcccc 35580
gccgaccacc ccacccagga aagaacctgc cgtctgcagg aacagccatt atttaacacc 35640
ccgtcccggc cacctgctct ggtctgcccc ttcaccgtat ggagctgtag gacacccagg 35700
gagggaccgg gggcagggtc tcactttgct tctacgaagt agtagcattt gagagaatga 35760
gtcaggaaag accgaattgg gtggtgccat ctcaagtgcc aagagagggt ttggcgccag 35820
agtttccaag ccctctggaa gcccagctcc ctaggggcat tctttcaggg ggccttcact 35880
gtccattctt gggtagcgaa gccatcagaa agacctctct gcctcctttg gacatgctgt 35940
taaagaccag tcattcaaag ctcaatcatt caagtgctct cttcctttct tttgatccta 36000
catcatgcaa gtacactccc aagtgaggtg acaaaaacgc acatttctgc agatgcccat 36060
gaaccaggac ctgtctctca agtctgccac ctgaggacag aatcctgtaa ccaccccagc 36120
tcccaatcag caggatgggg tgtcagggct attaaacaaa cagccacgtg gcccaacacg 36180
tctcatccgg cagctcgggg aaagcacact ttgacagcca ggaccctatc cgataaatgg 36240
tgcttcctct tcatgacaaa gaaaatgaac actcattcac tcaaaatatt cagcacctgc 36300
tgtgtgcttc actcttcctg gcacagggga tgcagaatga acagagagcc cctgccccac 36360
tgggaggggt gtttgtgggg agatggacca ggtaccagtc agtgaatata gcacaatggc 36420
aggtagagaa aagtgctaca gtcatctacc gtgagcgctg tgatgctctg cccagtttca 36480
ccacattaat ggagcaccca ctatatgctg gacacatacc atgcattttc tcatcctagc 36540
ggctgttgca aaatagacac gtccattgtg aagactgcgg gcggtagaat cgcagccccc 36600
aaaggtgtcc aggtcctaat ccccaaatcc tgtttgtgtg ttcccttgca tggcaggagg 36660
gatgttgcag atgggattca gttaaagatc ttgataccgg ggagatgatt ctggactttc 36720
ctatccagaa gggcccaaag taatttcaag tgtccttgta agagggaggc tggagggtca 36780
gggtgtgaga ttggaagatg ccaccctgca agcttcaaga tgggggaagg ggccatgagc 36840
caaggaatgt gggtgtctct cacgctggaa aaggtgagga gatggttctc gcctgcagcc 36900
tccagagaga cacagctgac agagtcagac tgggccactg acagggctct agcgtcccag 36960
attcccatgc ccccaccccc ctgcatggct cctgagcccc tctctgaggt gtctcctgga 37020
aaacctttgg gctcctgctc cctgcctgtt atccctgaat gttgctgcct tttgcagctt 37080
gtctgtccct ctctgaggct catgatgtca tctgcgtggg ccccccgcgt aattctaaca 37140
gcacccaggt aaagataaag gtgctttgtt gtttttctaa aaaggacttc gctgttgctt 37200
gagtcactca taacttgatc aattatttgg agctctaaaa gcttcccagt tgctaggaaa 37260
ccacaccttg gaggctggca aaccattgtt aaagttgttc ctaaagtgct gaacttctac 37320
agtgactttg gaaagtagct tggggaacaa aacaaagcaa acaaaacaga aaagactcgg 37380

gttccctttt ctttccgtcc agaagtgatc aaatcagcta aactcactgt ttcaagcttt 37440
cctctaacta gatgtgtttt ttttctgccgt tagccagaaa ggccctttcc ttcctatttt 37500
agtcacagaa taacctgtct ctttaggaca gtggggtcct gtgctgaatc ccacccccaa 37560
agatgggaag aactcgggtg ctgcgttttg ggaagccccg gctgagatga gtgtataact 37620
ggggctgaga gccgagagcc tggcccagca agccaggcag tgaggtgacc cggtatcccg 37680
ctccccattt ctgctggagc tggaagagcc ccaaggacac cagccccatc agagggagct 37740
gggcccagag ctgggtgggg cagccgacta ctccctgcag aacacctacc gcctgcctga 37800
cactggtggc tgctttgtgg acatgctttt gaaaacctcg tgagggtgat gttattgtcc 37860
ccactttatc caggggtacc tgggcacgga gcagggcggt aactcacctt ggatgtctgt 37920
ggcagcctcc aagatggtcc cagcaatccc caccaccttc ctgctgctcc ttccatgtag 37980
caggataggt ctgtgcgacc agtagaatat ggcagatgtg acgagatgtc actttctgagg 38040
tttgcttaca aaaggctgtg gctttcagtt tggctgctct ctctcggatc catcactctg 38100
ggtgaggaag ccatgccgcg agcagccccg tggggaagcc tgcacctcaa ggacctggag 38160
cctcaagcca actgccacat ggtgcacttg gaaacaggga tggctgccac ggcccacact 38220
cgcctgcggc tccggggaga cacccagcca gaacctcctg gtgaatccac tcctggattc 38280
cctaccccca cagagtataa gatgacgaat gtggctttag gctgttaaat tcgggataac 38340
cgatacagtg ccaacactgg taagtgacga ggctgcaatg aatcccaggt atcctgtggc 38400
tgcatctcca cctgcatggg ccatcctgta cctggcaggc cccatcctta gcctctttaa 38460
aagtccagga accttggaat gtttttctga catctcctag gctgctggat ttaccaggaa 38520
aagggacgga cccacacctt ttccctgttt gaatccattg catttcttcc aggtgtttca 38580
atcccttctg actcgatttg ctttaatgaa gcttgaggca agtgcaattt tctgcccatg 38640
gggctgtaac actgttcaaa gtcatcctgg agctaagcta agtggaaaaa accatatggc 38700
attgtcgcct tctcagcctg acctggcatg cactctcacc ctcatctgtc atgcctcctg 38760
cttttccacc tgggggggctg agaagtccgg ccatcgaaac cttggttcct gccagccacg 38820
ggagtttgga agctttatca gattcctgaa gcctcgtttc ctcatgggaa cagtgcaggt 38880
gaaagcacct tcctctcgga accgggggga agatgagagg aaattaaata gatgtatggc 38940
cccgcagcag gactggcgct ctccattgtg tctgaaattg gcaggttctt ggtctcactt 39000
acttcaagaa tgaaaccacg gaccctcgcg gtgactgtta cagttcttaa aggcggcgtg 39060
tctggagttt gttccttctg atattcagat gtgttccgcg ttttcctcct tctggtgggt 39120
tcctcctctc gctggttcag gagtgaagct gcagaccttc acggcgagtg tcacagctca 39180
taaacgcagt acagacccaa agagtgagca gcaattagat ctatcacaaa gaacaaaaga 39240
acaaagcctc cacaccacag aagacaaccc gagggtgtca acattactgg ctccagcagc 39300
ctgcttttat tcccttatct ggccccaccc acatcttgct gattggtcca ttttgcagag 39360
agctgattgg tctgttttac agagagctga ttggtccgtt ttgcagggtg ctgattggtg 39420
cgtttacaat ccctgagcta gacacaaaag ttctccaagt ccccactaga ttagctagac 39480
acagagcact gattggtgcc tttacaaacc ttgagccaga cacagagtgc tgattggtgt 39540
atttacagtc ccttagctag acacaaaggt tctccaaatc cccactagac tcaggagccc 39600
agctggcttc acctagtgga tcccccaacc gggctgcagg tggagctgcc cgccagtccc 39660
gcgccctgcg ccctcactcc tcagcccttg ggcggtcgat gggacccggc gtgccgcgga 39720
gcaggggcgg gcgcttgtcg gggaggctgg gccgcgcagg agcccacggc caggggagg 39780
ctcgggcatg gtgggctgca ggtccccagc cctgccctac ggggaggcag ctgacgcccg 39840
gcgagaattc gagcgcagtg ctggcactgc tgggggaccc ggcgaaccct ccacagctgc 39900
tggcccgggt gctaagcccc tcaccgcccg tggcgggcgg cgccgaccgg cttttccgag 39960
agcgaggcca gctgagcccc cgcccacccg gaactctcgc tggcccgcaa gcgcagcgcg 40020
cagccccgtt cccgcccgcg cctctccctg cacacctccc cacaagccga gggatccggc 40080
tccggcctcg gccagcccag agggatccgg ctccggcctc ggccagccca gagaggggc 40140
tcccacggtg cagcggcggg ctgaagggcc cctgaagcgc ggccagaggg ggcgccgaga 40200
gcgagcgggg ctgccagtac gctgtcacct ctcaccatca ctgcccttgc ttgctctacg 40260
gctgctcaga cccaaggggc aagtttctcc aggtatttcc aaaccaggga gggctttgtg 40320
tttagtttaa gtcaaggagg ctgtcccctt ttcgaggctg tttggatttc tttcagcaaa 40380
aactagcaga agttattacg caggaacttt ttgtagcata taaagaatga atcccgatag 40440
gtactgtgat ggttaccgtt aggtgtcaac ttgattggat tgagagatgc ctagaaagct 40500
agtgaagcat tgtttctggg tgtgtctgtg ggggtgttgc cagaggagac tgaaatttga 40560
gtcagtggac tgggagagga agacctctcc tccatgtggg tgggcaccat ccaatccgct 40620
gccagcgcgc ctccaactaa gcaggcagaa gaaggtggga gaagctggct tgctgagtct 40680
tccaccttca tcttcctcac atgctggatg cttccttcca ctcctcctgc ccttggacat 40740
cagactacag gttcttcggc ctttggactc tgggacttac accagtggtt tgccgaaatc 40800
tgtgctttct gccacagatt gaaggctgca ctcttggctt ccctactttt gaggcttttg 40860
gactcgggct gagccactac tggcttcctt cttccctagc ttgcagacag cctatcagga 40920

aactttgctt tgtgatcatg tgagtcaatt ctccctgata aacccccttt catagacaca 40980
tagatcctat aacttctgtc actccggaga accctaatat agttaccaat agaaacctta 41040
caactttccc tagattctgc aggctgggtg gagttcacca ccctttgctt tgaggtatca 41100
gtggaggaag aacagaggca ctggggtcag gacagaatgg ggccagaccc cagctccaga 41160
acctatgacc gcataggttt gagagacctg cgttccagcc tggcctctgc ctttactgtg 41220
tggccctggg caggttccct aacctctgaa gctcaagttc cttacctcta gaacaggagc 41280
aataatagga cctgcctcat gggtttatgt agaatacgat agagtgcaaa gtgtgcgaca 41340
caaggcccag cacacagtag gtgtcaatca gtggtcacac ttgtcactat cctataggat 41400
gggccttggg cacacctcct aacttttcca gcctcagttg gctcctctgt gagatcagga 41460
ttataaagat ctgcttttca gggtggctat gatttgttga gaaatgaagt atctatcaca 41520
ccatcattta cacaatcatt cattcatgga gccattcatt ccttcaatca tttattcaat 41580
cattcatgaa tttatggatt tagtccatca tccattcatt ctccagttct caaggctcct 41640
gtgctgggca tggtcaagac acgctccctt ccaatgggca gataggtaag cgggggcctg 41700
aggaagtgca gggactgtca tcagcaaacc ctgctccaga gtggagacag agcagagcat 41760
gttccaccca ggaacccctt ggtttggagg catctccatc tcctgccagc ccagctcagg 41820
ttctgccagt ccctgaaccc ttctcacact ctccacccca cccatcagga gtcatatggc 41880
agggctgtcg aggcgtgtga accagagcaa ctccatcttg aataggagat gggtaaaatg 41940
aggctgaaac ctactgggct gcattcccag acagttaagg catctaagtt acagggtgag 42000
ataggaggtt gacacaaaat acaagccata cagaccttgc tgataaaaca gtttgcagta 42060
aagaagccgg ccaaaatcca ccaaaaccaa gatggccatg agagtgacct ctggtcgtct 42120
ttgctgccac actcccacca ccgccatgac agtttacaaa tgccatggca atgtcaggaa 42180
gttaccctat atggtctaga aaggggaggc atgaataatc cgccccttgt ttagcatatc 42240
atcagaaata accataaaaa tgggcaacca gcagccctcg ggctgctctg tccatggagt 42300
agccattctt gtattccttt actttcctaa taaacttgct ttcacttttc tccatggact 42360
tgccctgaat tctttcttcc acgagatcca agagccctct cttggggtct ggattgggac 42420
ccctttccgg caacatctta gcttcctttc tctttaaatt aaattaaatg ctcatggttt 42480
ttaaagcccc tatccctttg ctctcacccc cagttgccat gcctcagagg aacggcctcc 42540
acccctttca gctgtttctt ctgctaatta ccttcatgtt ttggaataac atgttcaaat 42600
gctatttctg attttccagt ttgagataat atgtgtagac accatgaaag aaaaatcgtc 42660
tctcatatgg cccctctggc ctccttcaac gagactttaa tggacaagcc accttcccag 42720
tatcaccatc tcactgggtc ctgttaaatc agtgtccatc atttctgtgc tgtgtaaata 42780
ggattcactg ctgggccacg gaaggcacta tgattatgtt tctttctttt ttttttttt 42840
tggagactga gtctcactct gtcacccagg ctggagtgca gtagcatgat ctcagctcac 42900
tgcaacctcc gcctcccggg ttcaagcgat tctcctacct cggccttcca agtagctggg 42960
attacaggca cacaccacca cagagacagg gtttcatcac gttagccagt ctggtcacaa 43020
actcctgact ctggtgatcc acccgccttg gcctcccaaa gtgctgggat tacaggcgtg 43080
agccactgcg gccggcttat gattactttt tgtttatgct gtgcctaata attatctagt 43140
tttttgtttg tctagttttc tccacattga ccactaattc atttgtgaaa aatagaaaat 43200
ttttctcaat gtgttaaaac tatcagataa gctctcaatt tctgcttccc cagccccgta 43260
gatttgttgt ttggaacttc ccatcttcct gtgtcaatct ggactgttca ctctctaggc 43320
ccaccaccca gctgtcaccc tggggactcc ctcactgtca cccgggtgtt cctttgtgt 43380
ctctcctccg tggtatccct gtttctggga tcccagatgg attaactggg aaacaggcta 43440
ggctgctata acaaagagac cccccaatac agaagtttat gcaaaccaga attgtatttc 43500
tctctcaggt caccatctga gtaggtggcc taggctagtg tgctggctcc aggcaggttt 43560
cttctctctt gtcgctctct ctcttttgac ccacattttc tatcttgaga tctaaactag 43620
ttgctccagg tcctgaaacc acctttacaa aattataact gaggaaatta tgacagtgaa 43680
aaaaatcaga cctaaccgac tccatcttcc ttctaacctt taagctgtcc ttgttcattc 43740
ctggacatag gccaaactaa ctttggaaag gaattcagtt catggtttga ctgaaacaaa 43800
attgataaca gccctttccc aaaaagaccc cccttcttgc ctagggtcca gtctgccttt 43860
gcaggactaa caaattagct acaagattag aaatcacagt ttagaggtca tgcagcctct 43920
ggctctaaga gtctgaacct cggcatattg ctcctgggga taacatcatt attgtaaaac 43980
ctaagatcag tgcttgagat attttgcagg ccctgtgctc gatggattag ccaacaccac 44040
ccagaccagt aatctggctc aaccagttct accattgcac ctaggaagag aagacattaa 44100
gaaaacctca cttcaaccca ctatgattcc atctccaacc tgaccaatca gcactctcca 44160
cttcccaagc ccctacccac caaattgtct tcaaaaactg tgatcccggc cgggcgcagt 44220
aactcacgcc tgtaatccca gcactttggg aggccaagat gagcagatca cttgaggtca 44280
ggagtttgag accagcctgg cccatgcagt gaaaccccgt ctctactaaa aatacaaaaa 44340
ttagccagtc atggttgtgg gtgctgtaat cccagctact cgggaggctg aagcaggaga 44400
atcgccgaac ccaggaggtg gaggttgcat gagccaagat cgtgccactg cactccagcc 44460

tgggcgacag agtgagactc tgtctccaaa ttaattaatt aataaagtta taattgaaat 44520
taacactctg atccccgaat gcttggggag actgatttga gtaataataa aactccagcc 44580
tccctcacag ccagctctgc ataaattacc ttctccattg cagttccect gtcttgataa 44640
atcggctctg tccaggcagt gggcaaggtg aacccactgg gtggttacac tttcagcact 44700
gtgtctacag tctggccagc ctttaggatg ggagggcagg tgaaagacag gctttctccc 44760
tcttacagta gtgcctggag cagatcactt ccattgtcct cgctgttggc cccacctggc 44820
tctgggggac aatcctgtgt ccagataaca actccaatgc gacgtacgaa gggaaggtca 44880
gatgtcaggg acaaccagga gtccctgcct catgtgtttc cttctttcct ggtttcgtgt 44940
tagtggcgca caccctccag caacttcctg agaaattgcc tttatgtgaa cgggagccag 45000
tctgaaaact tcccaacata ggtccatacc tttgggtcat ctcccctcac ggtgacctgg 45060
gcttgctcct gtgacttgtt ttggcccatg agacgacagg aaatatgaca caagcagaga 45120
cttgagaggc aggtgtgcct tggggcttat cttactgaca ctgggggctc cagtgtcaga 45180
gacattggaa ccagagtgac tccatcttga atagggggctg ggtaaaatga ggctgagacc 45240
tgctgggctg tattcccagg aggttaggca ttcttagtca acaggagatt cacagatggg 45300
gaacaagtga acgatgttta ccaaacagac cccggactta acagatccag gagatgtcct 45360
gatgtcccga caccttaaga acaaaagcat tcttagttca aaaataagtt tggtggctgg 45420
gcgtggtggc tcacacctgt aatcccagca ctttgggagg ctgaggtggg cagatcacct 45480
gaggtcaggg gttcaagacc agcctagcca acatggtgaa atgcatctct attaaaaata 45540
caaaaaatta gcaaggtatg gtggctggca cctgtaatcc cagctacttg ggaggctgag 45600
gcaggagaag tgcttgaacc tgggaggcgg aggttgcagt gagtggagat cgcaccattg 45660
cactccagca ctccagcctg ggcgacagag caagactccg tctaaaaaaa aaaaaaaaaa 45720
gtttttgctt taaagattct tgcagaagac agtagttgca caaagatgaa caatccttag 45780
tcaccagccc tagtagcaga gcacacctcc cccaggattt tgtggctttg tcttatatat 45840
gaacaagcac tgtacctaag gcacatgtgt tcctcatctt gctcttggga acaccgtacg 45900
ctgcctatgg agtagcaatt ccttcttgcc tttactgtct taataaactt gctttcactt 45960
tactctgtgg actcgtcccg aattcttttct tgtgcaagaa ccaaaaaccc tctcttgagt 46020
ctggattgag accgctttct ggtaacaact ggggcgatca cagggctggc ctgggagagc 46080
ccgctgggtg agagagatgc ctgctgatgg catctgtgtt gtcctgagcc aaccagcaaa 46140
cctgtgagtg aggccgttct aaaacattca tcctcagctc tgcctcccag cccacacacg 46200
gaattgtaag aaatagtcaa tgtttttact cttttttttt taagacacta aattctgggg 46260
tgtttcatta cacaccaaaa gctaattggt acaaaaacctt gcatgggtaa aaagccttca 46320
tttccctctc acgtgtaatc aatagtttga aggggcatag ccagcattcg aggtccacga 46380
atttcttctt agagtttgaa agcagcgctc ccttcccttа ttgctgctgg gaagcctgaa 46440
gacatattct gattctccta agctatcgtt tacagtgttc tcagtctccc tgagttctga 46500
tcctcaaagg gatggccctg gaagtgggta ggctttccac cattgcgccg ggcactgtag 46560
tttcttttca tctgcaaagt ttgattcttc aggtcagaaa cattttgaat attttatttc 46620
ctccattttc tttgttgaac tcctcatatc tgtgtgctga ggcctctttg aggaggttag 46680
ctttgtttgt ttgtttattt atttatttat ttatttattt tgagatggag tctcactctg 46740
tcacccaggc tggagtgcag tggcatgatc tctgctcact acaagctctg ccttccgggt 46800
tcaagctatt ctcctgcctc agcctcccga gtagctggga ttacaggcac gtgccaccac 46860
gcccggctaa tttttgtatt tttagtagag acgggtttca ccatgttggc cagcttggtc 46920
tcaaactcct gacctcaagt gatcagcctg ccttggcctc ccaaagtgct gggattaaca 46980
ggtgtgagcc accacgccca gccgaggtta tctctttttc tcacctattt tctttgggat 47040
tctctttgat attttttgtt ctacttcctg agaaatttcc tcaaccgtat cttccaatct 47100
ttgcagagat ctttatttttt gcaatcatgc ttctaatttc catggtttcc ttagagtctc 47160
cacgtgtttc cttggctatg ttttccatat tctcttatct ctatgcagat atccattatc 47220
cgagttttga aaacggtatt ccacctggca ctgtagctaa aggttgcagc ttcttttcct 47280
tcctctttgc ttgtttgctt gtttttgtct tggtcttttc cacagcaggg ctgtccttga 47340
gtgtctggca ttctgggctg tctgcttcta tctcagcagc tgcctggaat ctctgtgtgc 47400
ctgtggcaga gtggggatgc gtgtcaaact gccagcttcc cagttggatg ctttggtgga 47460
gcagtacctt ttttttttt ttttttttga gatggagtct cactgtcgcc caggctggag 47520
tgcagtggtg taatcttggc tcactgcaac ttccacctcc catgttcaag ccattctcct 47580
ccctctcagc ctcccgagta gctatgattc taggtgcatg ccaccatgcc tggataattt 47640
ttgtgttttt agcaacgaca gggtttcgcc gtgttggcca ggctggtctc aaactcctga 47700
ccactggtaa tccgcccgcc ttggtctccc aaagtgctgg gattacaggg gtaagccacc 47760
gtgtctggcc gatggagcag tatctcgagg gacctccaag tgtccaaacc taggtctttt 47820
ctctccagct ggccaatttc ccagtgaggg ggcctttaac cttccccggg gtgtgggtgg 47880
gttgggtgtg acctgttctc agttggctgc ctgccgccag gtcttcctgg ggtcccggtc 47940
cccgcagcct ctcgcacaat ccctggggaa tgaacctccc gccttctgca ggggtgggag 48000

agggatgctt gctgaggaag acattcagaa aacctcaatt ttctccccac cctcagaggt 48060
ccctggtgcc tccaggccca gagcctttct ggggttctat gatgcaaacc agttgccttg 48120
tggttgcccc caccgcgggc ttgggctctg ctctctcagt ctctaagctg tttccactcc 48180
cccatgtgcc ttcttgcttc ccatcttcaa ggccacggtc cagctcccgt tctctttgtt 48240
ttgatccctt ctttgactct cgagcctgtg ggagcgaggg tgtttgatct gcagtgttga 48300
gcagggttct ctgggctcct gcaggggacc cagtgttgag cccacataca cctcgctcag 48360
taagtagtcc ctaaggactg tgcagggcac cagggacacg gtggtgagtg aatgatggtc 48420
cctggtctca catctgtgtc cgtgccagtg ccaggggggcc tgggctgtgg ggaaaggagg 48480
aatcgaggcg ctcagctcag gaggcttttc aggagcaaag ggtgcacagt tggtttcagg 48540
cagggaaagg tgtgagcagc ggctgggaaa acacacacca gaatgtcacc ttccgaggcc 48600
tgaggcccca ccaggtagcc agaaggtcca ggatacagtt tagtaattaa gaccttgggg 48660
ctctgcggac acgctgcctg gggtcagacc ctggccccaa cactcacagt ttgtggccat 48720
gagggactca cccacctctt cacgcttcag tttccttatt tataaaatgc gcggaaaaat 48780
cacagaatcc tcttgatcag gtggctgtgt ggatggatga gatgatgtgt acacagccct 48840
ccgtatagag cctggtgtgt cctctgtgaa tactaggaat gacatggagg gcaggtcagg 48900
gaggaagcca ggtcacccag gcttccccca ccccccgcca agccaatccc tggacctagc 48960
ccagtgggtg gaggtcaggt cgggcaggag ggtaggtcac ccaggggcgc ccccgccaag 49020
ccaagccccg gacctagccc agtgggtgtc aggaacgtca ggacccgcag ggctggggag 49080
agatgagctc cggagagatc cagggaggga aagcattgag catcatttct cccagccctc 49140
tcaaggccat ctctgcctgc cacagctctg agcgcagggc ctggtgtgga gtgggggccc 49200
atccatgtgt gcagaccccg ctggagcagg ccaagctgtc cctggccagg ggccagctcc 49260
ccaccactcc acaggctcca gccctcacac gcctggccct ctagaagccc tgctcactac 49320
ctctgccacc atctggctgc ccctctgtcc tctgcagccc cataaaccac caccagaaac 49380
tccccaggaa agagccctac ccttgccctg ctcaagaaag gcaatgaggt gccaccaggg 49440
ctacaggccc cgctttgctg cttccctgcg tgctgtggac ctcccaaggg cccttgcctg 49500
tgagcaggga aaatccctac ctaaaggggt ggctttgtca ccaggacgaa gtgcccaggg 49560
atatggggtc cctggcgctg ccatgacaaa tgactgcaaa ctgagaggct taaaagagca 49620
tatttattat cttccagtct ggaggtcaga ggtctgacac aggcctcact gggctgaaac 49680
caaggtgtgg gcagggctgg ttccttccgg aaggtcccag gaggatctgt ttccttgcct 49740
tttccagctt ctagaagctg cctgtgttcc taggctggtg gccccctcct ccatcttcac 49800
atccagcacc gctggtcgag tcttcatgtc atgccttgct gagccttttt ctgttgctac 49860
agctgtctcc aatacaacca ggaacggtcc tacactttca cggacttggg tgattagatg 49920
ggtcccacat ggataatcca ggctaaattc catcttcagg tccatgacct taatcacatc 49980
tgcaagatcc cagagctggc aattggtcaa tgttggctga agttgggttt ggggaggctg 50040
aggctgcagg atccatcacc accctgggcc tccgcacatg cctgccactt ccacccctgg 50100
ggtttgaatc ctgacccctc cacttattaa ttctactaac ttgggtatca gttaaattac 50160
ctcatctctc taggcctcct gttctccatg tgtacaatgg gtataacaat agcaaccccc 50220
atgctctgca gaaaggatag catgcagcca caatgagaca tcccttccct gtaggggccg 50280
ggtgacctcc agcacagcag gactgcgggg ttcagtcgag tcaccccaca gtactgagga 50340
ggagcggcat gccctcccac tatgtacaca cacagtcaga gaggttagtg agtggccagt 50400
gacgcacatc caggaagtgg ccacctcctc cctgtgctct gtcctggagc tgtggaactg 50460
agcgccaatg aagcccctgc cagcagaccc gcagaatagc acctgccttc ctcctgctga 50520
acctgaagcc cagaattccc tcctgaaagc cagaggcagg gcctggctcc aggcagggtg 50580
cagggaaacc caggcgtgcg tgggacactt acccacacac tctcctgttt agctcagcca 50640
ttgtggcagg tggggcgggg cctctctggc ttcctgaaga atccaggtac cctgtggttt 50700
ctctcccccg caggcttcca ttctctggga ccctcaggca cgtctgtggg ggagcactgg 50760
atgaggagtt atgtgctacc tgtgtgattg cgagctaatt ctttaacctc tctgaatcag 50820
tttcctcatc tgtatcatat tctcatggag agtgtgagtg tgtgtggaga tcatgatggg 50880
ctatgcacac ctgctgccag acaaaacaag acaactatca ttacggagtg tcatgtgcca 50940
agcactgtgt tttgcactat ccaagttaat cttgtccttc ctctcccacc acaaacctac 51000
aagataggtc gtcttattag ctccatttcc tcagtgatga attcagagct agggaaagtt 51060
aagagctgca ttaggagcta gagggagcaa acccttgacg tgaacccagc ttcccagtgc 51120
aaagcccatg ggcatcccca cccactgtat tgcctcctga tgggctccac ctccatgagg 51180
ttggactcac taaaaacaaa cagcattctg gacctcgttt ttgatgtaac atttgttatt 51240
ttattggaaa aagctggtat taacatattt atagttttat tcaacaattg ggtaatttgt 51300
gagacaccaa agaaaaaaag aatgcaccta tgagttacag agtccaaact gatcagggct 51360
gacaacttga ccaccatgta tcccacacca ccacccccac caccaccacc accaacagct 51420
tcgtcctcag agaagagcta aattaaaaac aaaaccaaaa aaaccccaac aacttcacaa 51480
tgactatgtg aacgcccgta cattcgagag taccaggaaa tgtaagcaga gccaggatgc 51540

aagtctgtga ctattacact ggtccttcca ccgtttctgg ttttgtccct cccctcgaca 51600
ccttttctta acacacattt gatcttcaca atttgctttg gacatcatta ttttccaggg 51660
atccagagac atacttttct tcaaagaaga tcttttcctt tctttttttct tcttttggt 51720
ttggtaccca aagtttaaaa cagaagtaag gaactaggaa ctgtattaag acctcctttt 51780
gaaggacatg taagatttgt actttgaaga tggtaaatgt taaatctatg atgtgacagt 51840
aacctctagc tgcaaataca agtttcttct cttaacattt gactcacaaa gggagatgga 51900
tcgtagcaaa tatccaagta atgggtaagg ccctcagagc tgggttttgg ctttaacacg 51960
ctatctatac aattgaccac attctcgaac ctctttacaa gggggaaaaa agtaatccaa 52020
tcctttgatt gttaaattta attagaatag atatgtacat ataatactgt ccaggtcaac 52080
tggattttat agtgatatat aaacaaacat tcttcattca cgttgattaa aaaaaaatca 52140
atgatccatt tccctgtgta aatgttatat ggccagaagt gagtacacat gcactttgtg 52200
cttttacaca cacaaaagta tactgtaatc cactgagaat aactcagctg ggtctgtttc 52260
ctgggttatg ttatatcttg taaaaaacaa aacaaacaaa ccaaaaaaaa ttggatttgg 52320
ctgggtctct gctccccact ttctgaatca aaatgccaaa ctacgtggct ctgctgcaag 52380
tccatgagca aagacgactc agaggggtgg gcaggtttgg tatcaccaga acaggagcat 52440
ctaccatgga aacaccacct tctctgtggc cctcattgtc agaggggcag agttcccgag 52500
ggatgtgtcc tctgggactg tgtggtcctt aaagtaaagg tatcctaaaa tggtcagaac 52560
atgcaatttc ctttcaaagg cagttcagtg cattggctca aggagggtgc aagcccagga 52620
tgaagtggag tcctggggag ggccacactc ctgaagccac caagcagagc gaggagctcc 52680
gggggtctt ctctgtcttc catcctgcgt ctcagttctc ctggaccctt gtggtcgcca 52740
agtctgccat gtcctacttc acaagacaga ggcattcctt tccaaacctt ccatttgaat 52800
gtcgctctaa catgagccgc ccacgagatg gactggagag aaccttgtgt gaagacccgg 52860
acgggttcct gacaccactg cacggcacct gccatgggtg acctgtgcat ccgctcagag 52920
cctcagcttc ttctcgtgga cagaggggat cctaagggtt cccgcttatc cccagcatcc 52980
agcccagagc ctggcacgct tcgagaagtt ggtgtttatg gaggaaatga atgaacaccc 53040
acctcacagt caggaggaag attccaaaaa atcgtgcatg ggagtcagtg tccgctccgg 53100
gaaggtgccc agtgaatgct gaagggtgag gcggcaggga gggctgcaga gggtaagtca 53160
actgcctcta gggctgagca ccacgcctgg tacacagcag aggctcagca aatggcaggt 53220
gaatcagatg aggacacagg cacacagaag acaaacccca ctccgctcct ctcgggaaac 53280
cgcaccagac ctgggacctg ggatgtttca agacattcag aagacctagg agaggagctt 53340
gcatctgcaa attctcagca tgaattcacc atccaaaaaa caacggtgag gtcatcagct 53400
ctgagacagg gcgaattgcc ccatttataa ccaaaagaac ccctggctgg ctcgtttgag 53460
actcatcatc tgtccctgga gccctgatct gggaccagtg aggcatgggg gagaagcagc 53520
tcccatcagc tccggtccct gcaacaggac gcactggagt gaaaaattaa cccggacgca 53580
taaaaagtcc tattgatgtg gtaaagaccg cagaccgaga caggaaggga cgtgaatgaa 53640
agaaccctga actgtaagac tccacagtca tgtccatttt atgatttgtg gcggtgaacg 53700
cttcctttcc tttttatttt tttaaacaga caattactgc caaacacaat tctggcctag 53760
gaaagctggg gcagggaggg ggcccaaact tcctgtgtcc acacactgcc acctctgcag 53820
ctgtcctcat cagtgctgtg actttcttcc cctccttgca ttgcggtcgt gaaggtcatg 53880
tcggggatga cttgcatgag gctgggtggc aggggccggg aactgcacat acctagtgca 53940
tgtcagagtt taccttgtcc tggaagatgt acaggttgtt ggtggcggca atggcaatga 54000
tgttctcagc cgggtgccag gccgtgtgca ggatcttctt ggtgaagtcc aagctgtcca 54060
cactgatgtc atcacgccgg cgcttgcccc ccacgcacac gcgccgtggc ttgagcacag 54120
cccggggctt gctgctttcc ctcgaggcct ccagggtcac gtcccgcttg gtgttccgat 54180
cgaacatgcg gaagaagttg ttgtaggccc cggtcatgat gacgctggtg ggagaaggag 54240
aggcatcagg tgaggaggag cacaagcccc ttctcgagaa ataagcccag gtgtgggggc 54300
tcacgcccat aatcccagga ctttgggagg ccccggtggg aggattgctt cagcctagga 54360
gttcaagacc agcctggaca acatagcgag actccacctc tacaaaaaaa aattaaaaat 54420
gaccgtagtc ccagctattc aggaggttga ggcaggagga tcacctgagc ctggggaggc 54480
agagactgca gtgagctgtg atttaccact gcagtctagc ctgggtgaca gagtaaaact 54540
ctgtcccaaa acaacaaaca aatgaaaaag caatcccaga gcctacagag ccatcactcc 54600
ctggaaaggg aaaccacacc tgcctctatt cccctgtatc ttggggtctc ttatacccca 54660
tcacatgtca gctgccacca acatgcagac tttggggcca ggagctaaga tccttttttat 54720
cctctgcccc aatcatcatg aaaatacccct cctggccagg cgtggtggct cacacctgta 54780
atcccagcac tttgggaagc caaggcgggc agatcacctg aggttaggag tttgagacca 54840
gcctggccaa catgatgaaa ccctgtctct actaaaaata caaaaattag ctggacatgg 54900
tggtgggcac ctgtaatccc agctactcag gaggctgagg caggagagtt gcttgaatcc 54960
aggaggcaag gttgtagtga gccaagatcg cgccactgca ctccagcctg ggcaataaga 55020
gcaaaactcc gtctcgaaaa aagaaaagaa aagaaaagaa aagaccctcc tgtaatttct 55080

gttagtccag gagagaaagc tctggccttc caaacagact gtgcaaccac agcaggaggt 55140
agacatggaa gcagctggga cagaaaacca ccaggccctg cttcctcca ggggcccgcc 55200
tgtcccgagg actcgggcga ataaacaaac atgcatcgcc atgagggtcg gggcgagcag 55260
tgctgacgtg caggggctcg ctgccctcac cgtcccgcta aagagaattc cacaccattt 55320
ctctgctctc cctgctccta aggagaaggg gttcctgccg gtgggaggaa gggtaagttc 55380
atccacatag taagaggcag gagagaaagc acagttctgt gggcctgcgc ccagagggag 55440
ggttttgggg tttggcctct gcagggccct ggagcagctg tgaggtgcaa ggtaactggg 55500
gaggaagagg atgttacgaa aagcgaccat aaagaagttg tcctgctgaa ggctcttcag 55560
ggaggaaatc ctgtttcctt ttggaagctg agctgggctt cctgagggtg acactgctct 55620
gtgacagaac caggtggatc cctgggtctc caggggttat gcacagagga caaaagctca 55680
cacgtgagct ccggcactgc agtcaggggc ttgggctctg gccagtccgc ctgggtccgg 55740
atcccagctc tgccctaccc tgacatgcgc actctcttag ttttctataa gaactgggat 55800
caacaaggga acgcagcccc tggtgagcgt gaatcagggg acataggctg aacaccccgc 55860
acagtgctgg gcttgaggca ctttgcttct tgttatttac caggtactcc tcgatcctaa 55920
tcgtccatgt gaccaaagcc gcacccgctt tatatgtgac tcagagcctg ggacaaggga 55980
agcagcatgc gaggggggcc gggtgaacga gtcccacaga agggccctgc ccaggcgcca 56040
caggctgact gggggaggcc acaccctccc cgcgccagca cagtggggag gcagcccccc 56100
ttgttcttca gggaacagtt cacagaatga acctgggtga tttcacacaa tgctggttct 56160
tctgatccat ttagtggaag gagagcgaca taattaaagt catttgcaac ccacacagtt 56220
cctgatcctc cccctgttcg ggtgtggaat tcagctgaca ctgctctggg ggagagggag 56280
gggcccctct catcctagag caccaggccg caccaccatg agagccattc ctcacattca 56340
ggcaggtttc ttaaattcca gaaggttccc ttaacaggga aaagggcagg ggagctgctt 56400
cacctctcct agcctcatct gcatgtggga ccacaacgcc ggctgctttc agccccagtg 56460
ctccgctcct ctttcctttg gggaaagaaa ggctcctgcc ctcgccaccc gtggcgcctg 56520
taatcccagc tacttgggag gctgcggcag gagaagtgtt tgactccagg aggcggaggt 56580
tgcagcgagc cgagatggcg ccactgcact ccctatccct gtgagcggtc aatcacagtg 56640
ctctcctcgc cccttggcct ggccaacttg ggcatgtgat atccctgtcc atgctgattg 56700
gtgcaaaaag tgggcaagtg acccacacag agccaatcag atccttccat gagagtttta 56760
cacagtccct gggaggggtt ggctggctga ggataatgga aaggggaatg atttgggggt 56820
catctctgaa gagggagagg attgaggcca gtgtgtaaga agaagcaggg aggagcggag 56880
atggggaagc atggagagaa gaaaacggcc ggagagaaaa aaacagccct agggatacag 56940
catgggcccc tgggggagcc ttgcccagga ctcctcagag caagtgtagg tgtgtttgac 57000
tcccgcaccc gagcacgctc ggctggcaca aatctctccc cccaggcagg acagatcaag 57060
gagaacttca tggaggaggg ggcgttgaag cctgttgtgg gagggcctgc gggattcctc 57120
cgggtaggta acctggggaa gggcaggcct gccaatagga tagtccacgc aggatcactg 57180
cggcacccct gcgcatacga actgggatct tgctgaaggc atcacacatg gtctcatccc 57240
caaagcggcc acgaaggcca gtttcctaga accactttcc agccgagctg acttaggggc 57300
ttcctcaagc acacggtggc gagggtagaa tgtggaccag agactgagag agtccagcag 57360
gagagctgct gcctctccac agcacgctgc ccatctgcgt tctgccggtg ctaggaagcc 57420
ccacggcagg ggccatcggg cagccaccac gcggcagcag cggcgctctg cgggccggac 57480
caactgtcag agggggtcac caggaataag agggcagcaa ctgggagcct gccaggattc 57540
gagtttgctt tggtctgcaa ggatgtggga acctgggggg aatccttgct tctctcctga 57600
gcctctgagg ccctggaagg caggacccag ggtctgttta tcttgggacc cctgaccatg 57660
gggcatgttg ttctcagaac agtgctcaaa gtgaactggg aggtggcaca cctctatgcc 57720
ttctagaaac acccacagcc actggacaca cctgaggtcc ccagtgaagg gctcagaaca 57780
ggttgagtca tagtagcccc ccagggcacc ccacccattg cgggaggcaa acaaggacag 57840
aggttccagc cgggcagagc agagccagct gtggagcgag agacagtgag gcgggtgggg 57900
gaccactagg acagggcgtg gctggcgtgt aaagtgcacg caggaagtgg cgggtgggag 57960
gcaggacagg agagcccagg cctagacacc actgtcagcc tggggatgct tggcggcttc 58020
tccagtcctg ggagcaggca tcacctggcc gcgggtgccc cctggtggca gcttgaagga 58080
aggacgggca gtgggtcgca gccagcgggg acctaccccg caaaacgcac ataaaagctg 58140
gaatcagctt gttacagctg caggtccctc tcgtccgatt tggatagacc ctcttgggac 58200
ccactgcacc agggaacccc aaatgcagct cagcagcatg ggaggagccc tgtctgctgg 58260
gggtgtctgg gatctatctg ggggcagggc ctagagagaa ggtcctgtgt cctgctgacc 58320
cagctgtgcc agcccagccc cgggtcgggc ctccatcctg gttttctaag catcctttag 58380
tggcccccctg tgatcaccat gccagcccct tgcctgagct cgggctctgc tgagagaccc 58440
ctgcgggctg ggggagctcc aggatgttgg caactctctc agaaatctcc ttccctggac 58500
acaatgactt ggcccccccac ttggatccag ccctgcttcc agggggctcc ctcctggtcc 58560
tgtctctgcc atgtccagcc ctccctcat cccacacacc aaccctgatg gtcccagccc 58620

tatgcccaga gccacagaga gtcatagtgc ggcagctctc acatgaccac cacagtgcca 58680
gccaccctca gaggtctggg agagcccaga cctgtcatgt tgactgatgc catccacagg 58740
gtggacagcc ccccaccagg tgatgcccaa gtcagggctg ctggctctca ctccaccctc 58800
cttgcccacc acactggaca ggactgcggg gaggaggtga cagtacatgg gctcctgaca 58860
ggtggtattc tgatctagca gccagtgtag tctttatctc cacactgtgt atcatctgga 58920
gacccacttc tgaatctggg cggctggaca aagaggcatc ctgtccttga ggaagcagct 58980
gtaggagaca ctgagaggcc acctctcagc ttctgagatc gaaactcaca gatggagagt 59040
aggcttgtcc atttcagaat cttacgcctg gcaggaccct gagcacaaag ctgatgtgat 59100
ttgggagggc agaagaggga ttttcccacc tccagaatgg aggaagaatc ccagagaaga 59160
gacacagagg taaggaccca cgggtagagg tcaggtcagg agcccctgcc acagcccct 59220
tgtccatgcc cgccctgggc tgagcccagg ctagggagac acaggaactc acttctggca 59280
ctgggggcgcc actttccagg caggtggggt tgagagacgg ctggacttgt ggattgtgac 59340
aagccctgga gggtgtagga gctccaaggt gcccctgagc ctcagcgggc ccgagtgggg 59400
tgtcaccagg aagagaggga gacagtagtg gtgatgggag gaggggtgag tagaccgatg 59460
gtcaaagcca caggggtgca gggctgtctc aggacataca ggtgtggatg gatgatggtg 59520
accgggtgct gggctatagc cctcgccccc gacaagctgg gagaggggag cacatcaccg 59580
gggaccctga attaactaaa catgttccca ctactgccca tagggcaggc gccaacgtgg 59640
ctgtgaggtt cagtgacaga agatgaagaa tttcatgtcc tgccccttga atctgagcgc 59700
tgagagttgg acctgctctg gaaagcgtgg gcttcaccca gacacctgga cccattgagg 59760
ctgagtagcc ccatgctgcg ggtcaccgga atcccagccc agacccctgc agggcagaaa 59820
ccctccctac ggtgaggtga ggtgcgggct gaggtcaggg cttacctgtc gctcccgttc 59880
caggcacatt caaacttgtc gaaaatgcag tcgttctcgt acagggaaca gagcttgctc 59940
cgaaggtagt catggacctg gtgggataag ggatgaggtg agtggacggg gcgtcccgac 60000
catcctggcc cttccacaag aagggtctca aagagcagcg agggtctgca tgccctgaca 60060
tggcccagcg cagacctgct catctcagcg agggtctgaa tgctctgaca cagcccgaca 60120
cagccctgct catctcatcg ggaggcccat gaccaggcac acggctgacc ccgaccgtga 60180
cacagcccaa tacagccctg ctcatcttat cgggggggcc acgaccaggc acacggctga 60240
cctccaccgt gacacagccc aatacagccc tgctcatctt atcgggggcc cacgaccagg 60300
cacacggctg acctctgagc tgtggccaga gacgggagta gcaccaggga gcagagtcca 60360
ctgtgacaaa cagaacgctc acagcacagg ggccaggacc acctggggtc cagatcctgg 60420
ctctgccctc ccagctgtgt gatcttggac aagtgatttc atctctctga gcctcagttt 60480
cttcaaactt aaagtgataa gagggcgaac ctcacagtgg ttctggcatt aagtgcaggg 60540
caggcctttt gagggtaatg atctccgaga acacagggtg gcggccatca agaaagggga 60600
agagggaacc cacagaggtt tctcatccat ccacagcacg aggcactgag gctcacaggg 60660
tcacacagcc tctaagaggc tgaccaggac ccttccccag gattcccgta ccccacagga 60720
gagagggtga caccccagga cctgcatcca cctgactgca gccggtccta caagccctgt 60780
acccggaggg ccgcatgagt gccagccctc tgctgaggct ggagctgcca ggacaacagg 60840
gcacgggctc tgccctctgg gtgcggaagg gaaggtaact gggtgccatg gagcctggtt 60900
tcaactccca ggagttgaaa ttaacaactc caacgaacac agaaaatgtg aagggtggtg 60960
tgaacgtcaa tttcgtgtgt cagcttggct gggccaaggt accagatact tggtgtttct 61020
gtgaagacat tttgtagggg aggttaacat ttaagctgat atactgagta aagcaggctg 61080
ccctccatca tgtgggggggc cttgtccaat cagtcgaagg ccctgataga acaaagacag 61140
acctccctga ggaccggggg attccgctaa cagatgcgtg tgggctcgca tagtcactct 61200
cccctgggtc accagcctgc ccgcctgctc tgcagattta gaacatgcca cgtgagctaa 61260
ttccttcaaa tctcccttct ttctcagtct ctatccccat cctgttggtt ctgtttctct 61320
gagccatctg actaatatgg tggcaaggac tacacaggaa aacagagtgg cactagggga 61380
agcaggattg ggaggaaggg cctgctattc actgcacaca ggatggttag acaaggctac 61440
ctgactcccg cctggggccc tgctcatccc tccccagccc ccgtgtcctc ctggtgaccc 61500
ttgccttccc cttgggcccc tgggctgcag ccccaccaag ctgagggtca ctgtgttcct 61560
gtctgtctgg gaccctcgag ggtgaggccg aggttcttcc ttctctctgg gtccagggtc 61620
tagcactgtg ttgggcacag ggtactggct cagtatgtgt aacaactaaa atcaaggggg 61680
atgaatggca gcaggtttta gtgttcgtgt cccacccccg gctctgcatt tttgtgtgtc 61740
ctccttactc agctcccagc agccttggtg atgcacgcgt ccaaatgcgc atgctgttct 61800
ctctgccgtg cactaagcat ctcctgccct ctagtggcag ggtgaggaag agctctggag 61860
aaagactgtc aatctaaggt ccgctgcttg ctggctctgt cacagatttg ccatttaact 61920
caattgtctt ccaataaaat gggatgggga gaataccggc cccacctccc aggacagctg 61980
ggaggcttag atgaggcaaa gcatactatg gacttacctc agtgtggctc attcaacaac 62040
cgtccccagc cccttcaccc ttgccttctt caaccacaga ggctgttaaa ccattaatct 62100
tcccagccac tatggtggct gccagggacc cattctagac aatgaggtgt aagaagtctt 62160

ctgagaagca ttctggggaa gattttacaa tcacagtgaa acacgcaaca gaccccaggg 62220
gtgagatcta gtgcatttcc tatttcctgc ctggaatgta ggcaaagtgc ttggaggtgt 62280
ggcagccact ttgccaccat gaggtaccac atgtgaaggt ggaaaaccaa cgcccaaatg 62340
atggtggact ggatcacaga aggaaccact gtggcgctgc cggggtcatg gagccccccc 62400
accttcctgg actgccctct ccagactccc tgttgtggga gaaaaatagc cccgatttgt 62460
taagcaccgc acaccagccg aacgcagtcc tgctacacac aatttctagc acatcacaag 62520
ccaccaaaaa atgttggttg ttgtgctgct tttatgtagt tattgctatt attttagcat 62580
caataggtca gctgggccag atgccaagtg cccagcttac ggcctcacgc tccagaacag 62640
tcattcccct ccttcgtggg gtccttcatg agaatttcag aaagctgtgt ccccctcccc 62700
ctgaaaatgc tcatagaaca cattccacgt agtttcaggg gttcatgggc ctttgctgac 62760
ctcagggatg gaggtaagtg accggacaag gggccggagg gggttggaaa tgttcatttc 62820
cctgagctgc ctccagaggg agccagagag ttgctcctgc aggcttgagc cgggcttatc 62880
ctgtccctct cctagtctgg ggtctctgcc attcccacag ggacagagga gggctgagtc 62940
ctgcctgggg tgtgcagtgt gaggaggggg cctgagggag gagctggggc agtcgcaaga 63000
actcaagtaa acacacgtgg ctgacaaaca ggcggtcccc atagcggtcc ttgtgggtgt 63060
gtggcttcca ggagcgtgag ctgttcctct ggggcccagg tcagatcaga gcagcccacc 63120
ctctgagctc tgcaccagtt cccagcgagg cctgcgccac ccaggacgct ttcctactag 63180
gtcttcgcac ggaggaaagg aaagcgtagg gctctgagcc tgggctctga atctgtcacc 63240
gcccagccga gagacacctc cctcatctcc ctacctctca gtgcccgagt tccctttcca 63300
ccaaatagga tgatggctcc ttctcaccgg gacagtaaag aatggaggtg acagatgtca 63360
agcacgtggc accccatgtg tgaggagtga cacggtgagg aaccggctgg ggtttctggt 63420
ggggtgtttt cccacccatg tttgcacggc cagatcttac ccatccttca aggcccagca 63480
caaatgccac ctcctccaag aagcctccct gatctcccca cctaactgga accagctttg 63540
aagccatccc acatttcttt caagtccaca ttttatttta acatcacatg agttttgttg 63600
tctacccccc aacatatctg aatggatgat aatagcaaga taataaataa tactgtccaa 63660
gaagggccca ggcttcacct agactccagg cttataggcc tgtgctgagg atctggaaag 63720
gaaacatgtg ctcagttcac ctctgcgtgc actgcttacg gatgcaccct gggagggcaa 63780
cgcttaaccc tctttaatgg tacagctccg atgcaggctc ccagggagtg gggagggaag 63840
ccatcgtcta catgctccac tggggcccag cctgacgctc atctctgttc ccacatctct 63900
ttcctccctg gaattgaggc ccctgggccc tggactgcag gccttgctca gggacatcca 63960
aagctccctg tgaggcaaga ctgagcactc ctgccctctg ccaagctacg tgtggggatg 64020
agttgctgga tttcttcagc ctccgtgtct tcacctacat actgggcata tgaaagccac 64080
gcctggctag gaaggccccc tccatgggga tctcagcata gggaaaaaag acccgggatt 64140
gggggtctcc tgctgtggtg cccacctggt aggtctctat gggtcttgcc tccatgttca 64200
ggtcccagac cttgactgta aggtagtccc gggtgagcat gtagcggccg ctgtggctga 64260
acttcacgtc ggacacggag gagatgattt ccgagaagaa tgagcggtta ctggggtcct 64320
caggctcttc aaagactgtg gagacagaga agcaatggcc gtcactgcgc tgctcccgcc 64380
catggggttg gcacgacgga tgactctgtt ttgcacctgg gcccatgctc tgtttattca 64440
gtcattcatt cattcctcaa acctagaact aaacacttac cagggaacaa gcctggtaag 64500
attctgtgct gtgcagaaat agccattact atccccattc atagccaaga aaatggagga 64560
gcagctggct aaaacccatg cctggggagt cagtgctgga agaaggtggt agagtcggga 64620
tttgaagcca aagctatact tgatgtgctg ccctctactg ccccctactg gtgatgctga 64680
atgccagatt gcaaaggtgt cctgacattc tcaaatgtgc tggacaatgt cctctaccag 64740
actgggcact gcttgtggct ctggcactaa ctggccctgg gaccttgggc cagtagttat 64800
ctctgttcat ggtgtcctgg gcaggcatgg ctgagcccta cttataggtg ggttttgttt 64860
cattgccatg gggtttgtaa caaaatcaaa tctgactgcc acagagttgg tacagtctcc 64920
tgtatgccat ccactgacca tctccttttt tcttggagca cctgctctgc acaggaagtt 64980
gcttatctgg tccctaaagg caacggtttg ctgcccacct ggacaattat agactgggac 65040
agagtgggca atgaggactc cccagacagt gtcccctccc cactgccccc ataacctggc 65100
cttcctgagc tgcagaagtt cccaaagtgg gacaaggtct ctaagctcat ttctgctggt 65160
atctgtctca ctggtagcag tgaccctggg gctgaagtag cgtggggact aggacccctg 65220
ttgggccagg cactaactcc ctaggacctg ggccacagag agacccaggg aacgatggga 65280
agagaacagg gtggtggagt tgacagccat tcaccagcca atctggaaat acctaaacct 65340
atgacaacct gtccagctgt ccccatggaa accggctgag tggcagcctt gcctgctggc 65400
ctgctctggt ctctacccca gaaacaggga agcagagctg atcacagcac cctgaggctc 65460
ccgagaggcc cctccctgct ggcccctctc agccacaagg gctcgtgtcc tgcctcccag 65520
gtctgtccca gccccttatc ccctcatggg aaatggcagc accccattct ccctcctgtc 65580
caaggtggac actgtacctg gccactgcct ctccagaaat acaactcctg tccagctccc 65640
ctctgcaggg cccactggat gccctgatgg actcattcat tcattcattc atttattcac 65700

tcagtcaacc aatgcaaaca gcacctactc tgtgccaggc cctgggatga tgcagtgacc 65760
agcataaagc cactgttcct gtgcagccaa cattctcctt gcagatacag acaataaaac 65820
taagtggcgc cctccaccgc cagggagcag caagtgctgg ggaggaagag ccagcaggca 65880
gataggtgcs ctggtagttt aggtgaacag ggcctccgtg ggaggtgctg tttgaccaag 65940
gtctgagtga agttggggag agattggtgg cgatgcctgg agaggggtgt gctgggcaga 66000
agggacggca cacgcaaggg ccctgaggca gggacacgcc aagggccgtg caggcagaga 66060
aaggagtgag caggggggaac agggacgcaa ggggccaccc acgcaggatc tcatgggcca 66120
cggagagggg tgaggaagtc accattaggg aaacgggagc cactgagggt tcagagacag 66180
ggaggatgaa atctgagtca aggaggtagg gaggggcagg caggaggctg gagtcccctc 66240
ccgacggagg aaggaagttc cagcacccgg tgggccaggg ctgcagagcc caggtgtcac 66300
catcctgatc ccaggccact gcatgcaggg aggtgagggg aggtgagagg agggggaggg 66360
aagctgctta cagggcagga aaaggctccc acgtttatgt tcatgcctct ggtatccccc 66420
caaatcagag caggcagcag gaggaaggtc ccccccacct ctcttgagag caaggcctcc 66480
aaacacctgt ctctgaagca gacccctgaa gctcaggaga tctatatccc tgtccttagc 66540
ctctccacag cctctggtga gcagtggcag ccttcaaccc cacacctctg ccctcaaagc 66600
catggcactg caccgggtat cctgagggca ggtggtcgct ccttccagtc ctggcgtcaa 66660
gatcaggaag gagggaggga gcctagttta tccactaata agtggtgtca tattttggaa 66720
ttagaatgct atcaggctgg caggtgttta acgaaactca aaactgacag taggcgcctc 66780
ccttgccagg agcccaaccc acccgtccac acgcccccag tcctcaggca tgcgtgcgac 66840
ctcaatctcc ctccatcgtt cacaggggcg cccttctcag tggcccgagc tttcctcttt 66900
ccacaagggc acattgaagg gaggaccaga aagaaaatca gcaccatggg aaggaagggt 66960
gaggccgctc gcagagggag ctgggagaga acaggctgcc agggtctcaa tccccgagcc 67020
cctcgtgtgg ccttgaacgc cttcacctcc ctgagcgctg gccttgctcc tccctacaac 67080
aggagccaag ggcgctcttc aggaaggggc aggaggaggg ggttgcagcc aggcctggca 67140
ccaccctggg tacaggcagg gccgcaggca gcaaagatgc tgtacagagt gagcccagtg 67200
cacgtgctgc ag 67212

<210> 2
<211> 3688
<212> DNA
<213> Homo sapiens

<400> 2
gtgcagaagg ccgcgctagc cggctcttca gcagcgagtg cagattgctc ccccgcggcc 60
gcagatctcc cgtttgcgcc gcgttcagct gctcccgaac aacttttctg ccggcccaga 120
ggccccaggg cgtcgcagcg ccgcgtgcgg cccactcacg ggccggcagg atggactcca 180
acactgctcc gctgggcccc tcctgcccac agcccccgcc agcaccgcag ccccaggcgc 240
gttcccgact caatgccaca gcctcgttgg agcaggagag gagcgaaagg ccccgagcac 300
ccggacccca ggctggccct ggccctggtg ttagagacgc agcggccccc gctgaacccc 360
aggcccagca taccaggagc cgggaaagag cagacggcac cgggcctaca aagggagaca 420
tggaaatccc ctttgaagaa gtcctggaga gggccaaggc cggggacccc aaggcacaga 480
ctgaggtggg gaagcactac ctgcagttgg ccggcgacac ggatgaagaa ctcaacagct 540
gcaccgctgt ggactggctg gtcctcgccg cgaagcaggg ccgtcgcgag gctgtgaagc 600
tgcttcgccg gtgcttggcg gacagaagag gcatcacgtc cgagaacgaa cgggaggtga 660
ggcagctctc ctccgagacc gacctggaga gggccgtgcg caaggcagcc ctggtcatgt 720
actggaagct caaccccaag aagaagaagc aggtggccgt ggcggagctg ctggagaatg 780
tcggccaggt caacgagcac gatggagggg cgcagccagg ccccgtgccc aagtccctgc 840
agaagcagag gcggatgctg gagcgcctgg tcagcagcga gtccaagaac tacatcgcgc 900
tggatgactt tgtggagatc actaagaagt acgccaaggg cgtcatcccc agcagcctgt 960
tcctgcagga cgacgaagat gatgacgagc tggcggggaa gagccctgag gacctgccac 1020
tgcgtctgaa ggtggtcaag tacccccctgc acgccatcat ggagatcaag gagtacctga 1080
ttgacatggc ctccagggca ggcatgcact ggctgtccac catcatcccc acgcaccaca 1140
tcaacgcgct catcttcttc ttcatcatca gcaacctcac catcgacttc ttcgccttct 1200
tcatcccgct ggtcatcttc tacctgtcct tcatctccat ggtgatctgc accctcaagg 1260
tgttccagga cagcaaggcc tgggagaact tccgcaccct caccgacctg ctgctgcgct 1320
tcgagcccaa cctggatgtg gagcaggccg aggttaactt cggctggaac cacctggagc 1380
cctatgccca tttcctgctc tctgtcttct tcgtcatctt ctccttcccc atcgccagca 1440
aggactgcat cccctgctcg gagctggctg tcatcaccgg cttctttacc gtgaccagct 1500

acctgagcct gagcacccat gcagagccct acacgcgcag ggccctggcc accgaggtca 1560
ccgccggcct gctatcgctg ctgccctcca tgcccttgaa ttggccctac ctgaaggtcc 1620
ttggccagac cttcatcacc gtgcctgtcg gccacctggt cgtcctcaat gtcagcgtcc 1680
cgtgcctgct ctatgtctac ctgctctatc tcttcttccg catggcacag ctgaggaatt 1740
tcaagggcac ctactgctac cttgtgccct acctggtgtg cttcatgtgg tgtgagctct 1800
ccgtggtcat cctgctggag tccaccggcc tggggctgct ccgcgcctcc atcggctact 1860
tcctcttcct ctttgccctc cccatcctgg tggccggcct ggccctggtg ggcgtgctgc 1920
agttcgcccg gtggttcacg tctctggagc tcaccaagat cgcagtcacc gtggcggtct 1980
gtagtgtgcc cctgctgttg cactggtgga ccaaggccag cttctctgtg gtggggatgg 2040
tgaagtccct gacgcggagc tccatggtca agctcatcct ggtgtggctc acggccatcg 2100
tgctgttctg ctggttctat gtgtaccgct cagagggcat gaaggtctac aactccacac 2160
tgacctggca gcagtatggt gcgctgtgcg ggccacgcgc ctggaaggag accaacatgg 2220
cgcgcaccca gatcctctgc agccacctgg agggccacag ggtcacgtgg accggccgct 2280
tcaagtacgt ccgcgtgact gacatcgaca acagcgccga gtctgccatc aacatgctcc 2340
cgttcttcat cggcgactgg atgcgctgcc tctacggcga ggcctaccct gcctgcagcc 2400
ctggcaacac ctccacggcc gaggaggagc tctgtcgcct taagctgctg gccaagcacc 2460
cctgccacat caagaagttc gaccgctaca agtttgagat taccgtgggc atgccattca 2520
gcagcggcgc tgacggctcg cgcagccgcg aggaggacga cgtcaccaag gacatcgtgc 2580
tgcgggccag cagcgagttc aaaagcgtgc tgctcagcct gcgccagggc agcctcatcg 2640
agttcagcac catcctggag ggccgcctgg gcagcaagtg gcctgtcttc gagctcaagg 2700
ccatcagctg cctcaactgc atggcccagc tctcgcccac caggcggcac gtgaagatcg 2760
agcacgactg gcgcagcacc gtgcatggcg ccgtgaagtt cgccttcgac ttctttttct 2820
tcccattcct gtcggcggcc tgaggatggt ccgccacgag gagcttccag tgcatgttgc 2880
catgaggccc ttccccagtg tggccccagc ccgacaggca tgcaccagtg ccgtctgtgc 2940
ccacgtgtgc agactgtggc tgcagagacc ttgcgaccat gtgtagattg catggacccc 3000
gacaaaggga aggctgctgt gtagctctgt ccactctgaa taccaagtgt gttgggaatt 3060
gcatgccatc tccaccctga gcctgacctt tctgagtgac atgggtgtgc caggctagac 3120
taggaggttc cggtgtctgg aaaagcactt tacagatgag attcccctct ctcccccacc 3180
ttcaagcacc ctgttccctc tttctttctt ttgtgttgga tttgtttaaa aaccaaataa 3240
gcatctgtgt aacctccaca gtagcatttc ttatttgttt ggtcactgct acaccttagc 3300
agctcttccc ctttcctggg ggatgtgcac ggcagcttga gcctgtcacg tggtcaaggc 3360
ccggccccat cagaggctgg gggaggcggc acattggcag tgtgtcacac tgagctgggc 3420
accacaggct gcctcatgac cctcctctcc agcaggtagt gggtgaatgt gtgaaggtct 3480
tgcctgaatc catcaggact tgggaaacag agaaccctgt gggggtggct gtgggggagg 3540
tccctgccag tgtttagaag agcctgactg tgttcagtgc cttggagcag aaagccaggg 3600
tcctgagtgg ctgaaataaa agcctctggt ggaacctgca aaaaaaaaaa aaaaaaaaaa 3660
aaaaaaaaga aaaaagaaaa agaaataa 3688

<210> 3
<211> 3512
<212> DNA
<213> Mus musculus

<400> 3
cggtttcgga gcaacttcgc ggcgggctgg gaggcccagc actgcggccg gcaagatgaa 60
ctcaggcacc ccacctccga gcccctctgg cccacctcct ccacccgcac cacagcccca 120
ggcccgggcc cggctcaatg ccaccgcctc actggagcag gacaagattg aaccgcctcg 180
tgctcccaga cctcaggctg accccagtgc tggacgaagt gctggggaag cagccgctcc 240
ggagcctcgg gcccctcaaa ccggcagccg ggaagaaacg gacagagctg gtcccatgaa 300
ggcagatgtg gagatcccct ttgaagaagt cctggagaaa gccaaggctg gagaccccaa 360
agcacagact gaggtgggca aacactacct acgccttgcc aacgatgcag atgaagaact 420
caacagctgc tcagccgtag cctggctaat cctggcagcc aagcagggca ggcgggaggc 480
cgtgaagctg ctgaggcggt gcctagctga ccggaaaggc atcacttctg agaacgaggc 540
tgaggtgaag cagctatcct ctgagaccga cctggaaagg gctgtgcgca aggctgccct 600
ggtcatgtac tggaaactca accccaagaa gaagaagcag gtggctgtgt ccgagctgct 660
ggagaatgtt ggacaggtca acgagcagga tggaggggtg cagccaggcc cagtccccaa 720
gtccctgcag aagcagaggc gcatgctgga gcgcctcgtc agcagtgaat ccaagaacta 780
cattgctctg gacgattttg tggagctcac caagaagtac gccaagggca tcattcccac 840
caacctgttc ctgcaggatg aggatgaaga tgaggacgag ctggcaggga agagccccga 900

```
ggacctgcca ctacgccaga aggtggtgaa gtacccttta cacgccatca tggagatcaa     960
agagtacctg attgacgtag cctccaaggc cggcatgcac tggctctcca ccattgtacc    1020
cacccatcac atcaacgccc tcatcttctt cttcatcatc agcaacctaa ccatcgactt    1080
cttcgccttc ttcatccccc tggtggtctt ctatctgtcc tttgtgtcca tggtcatctg    1140
cacgctcaag gtgttccagg acagcaaggc ctgggagaac ttccgtactc tcaccgacct    1200
gctgctgcgc ttcgagccca acctagacgt ggagcaggcc gaagtgaact tcggctggaa    1260
ccacctggag ccctacatcc acttcctact gtcagtcgtc tttgtcatct tctccttccc    1320
gctggccagc aaggactgca tcccctgctc ggagctggcc gtcatctcca ccttcttcac    1380
ggtgaccagc tacatgagcc tgagcagctc tgctgagccc tataccaggc gtgccctggt    1440
caccgaggtg gctgccggct tgctgtccct tctgcccacc gtgcctgtgg actggcgctt    1500
cctgaaagta ctcggccaga ctttcttcac tgtgcccgtt ggccacttca tcatcctcaa    1560
cgtcagcctc ccctgcctgc tctatgtcta tctcttttac ctcttcttcc gcatggccca    1620
gctgaggaac ttcaagggca cttattgcta cctggtgccc tacctggtgt gcttcatgtg    1680
gtgtgaactg tccgtggtca tcctgctcca gtctaccggc ctgggcttgg tccgggcctc    1740
catcggctac ttcctcttcc tctttgccct ccccatcctg gtggctggcc tcgccttgat    1800
gggcacggtg cagtttgccc gatggttcct gtcgctggac ctcaccaaga tcatggtcac    1860
cacggtgatc tgcggcgtac ccctgctttt ccgttggtgg accaaggcca acttctcagt    1920
gatggggatg gtcaagtccc tgacgaagag ctccatggtg aagctcattc tggtgtggct    1980
aacggccatc ctgctcttct gctggttcta cgtgtaccgc tcagaaggca tgaaggtcta    2040
caactccaca ctcacctggc agcaatatgg cttcctatgt gggccccggg cctggaagga    2100
aactaacatg gcccggaccc agatcctgtg cagccacctg gagggccaca gggtcacgtg    2160
gacaggccgc ttcaagtatg tccgagtgac cgagatcgac aacagtgctg agtcggccat    2220
caacatgctc ccgttcttcc tgggcgattg gatgcgctgc ctgtatggcg aggcctaccc    2280
atcttgtagc tctggtaaca cgtccacggc agaggaggag ctctgccgtc tcaagcagct    2340
ggccaagcac ccctgccaca tcaagaagtt tgaccgctac aaatttgaga tcacagtggg    2400
catgcccttt ggcaccaacg gcaaccgcgg ccatgaagag gacgacatca ccaaggacat    2460
tgttctccgt gccagcagcg agttcaagga cgtgctgctg aacctgcgcc aggggagcct    2520
catagagttc agcaccatcc tcgagggccg cctgggtagc aagtggcccg tcttcgagct    2580
caaggccatc agctgcctca actgcatgac gcagctgtca cctgcccgga ggcacgtgaa    2640
gatcgaacag gactggcgta gcacagtgca cggtgccctc aagtttgcct tcgacttctt    2700
cttcttccca ttcctgtctg ccgcctgagg agcgtccgcc gctggaggag gctttggtgc    2760
atgttgctgt gaagtccttc cgtgtggcca cccagccagc tggagcagca ctgtgccgtg    2820
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtatgtgta ttctgctctc gtgtggttag    2880
atcccaggct ctctgctcac ctgtagatcg cagatcctgc tggagggtgg ttctctttag    2940
cactgtccac tttgaatgcc gagtgtcata agaaattgca tgctatcttc actcacaatc    3000
ctgcccttcc ctcacagagc tggaactcca agcctggccc caaagaccct ccatcacgaa    3060
gagcacttta cagataagaa tcttcctccc cagttcttca tgcctccttc ctgccctttc    3120
cttactttgt gttggatttg ttttccaaat atgtgtgtat gtagacttca tggtagtgtt    3180
tcttatttat ttggttgctg ctaagccttg acagtggttc accttcctgg gctgtcccca    3240
gtggtcacgt cctgcctggc ttcctacttg ggtatagcat gtccaaactg ggctctgaac    3300
actacagcct gccttggagc tggcctatct ctgggggtgt tagagagttt gtggggatct    3360
cttctgagta ccccaggact tgggaaataa tgcaagagta tccctttcag tacataggta    3420
agcttggcta tgttcaatat ggcagagcca aaagccaagt tctaaggcag ccgaaataaa    3480
aagctttgaa acctaaaaaa aaaaaaaaaa aa                                  3512
```

<210> 4
<211> 890
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Asp Ser Asn Thr Ala Pro Leu Gly Pro Ser Cys Pro Gln Pro Pro
 1               5                  10                  15
Pro Ala Pro Gln Pro Gln Ala Arg Ser Arg Leu Asn Ala Thr Ala Ser
            20                  25                  30
Leu Glu Gln Glu Arg Ser Glu Arg Pro Arg Ala Pro Gly Pro Gln Ala
        35                  40                  45
Gly Pro Gly Pro Gly Val Arg Asp Ala Ala Ala Pro Ala Glu Pro Gln
    50                  55                  60
```

```
Ala Gln His Thr Arg Ser Arg Glu Arg Ala Asp Gly Thr Gly Pro Thr
65                  70                  75                  80
Lys Gly Asp Met Glu Ile Pro Phe Glu Glu Val Leu Glu Arg Ala Lys
                85                  90                  95
Ala Gly Asp Pro Lys Ala Gln Thr Glu Val Gly Lys His Tyr Leu Gln
            100                 105                 110
Leu Ala Gly Asp Thr Asp Glu Glu Leu Asn Ser Cys Thr Ala Val Asp
        115                 120                 125
Trp Leu Val Leu Ala Ala Lys Gln Gly Arg Arg Glu Ala Val Lys Leu
    130                 135                 140
Leu Arg Arg Cys Leu Ala Asp Arg Arg Gly Ile Thr Ser Glu Asn Glu
145                 150                 155                 160
Arg Glu Val Arg Gln Leu Ser Ser Glu Thr Asp Leu Glu Arg Ala Val
                165                 170                 175
Arg Lys Ala Ala Leu Val Met Tyr Trp Lys Leu Asn Pro Lys Lys Lys
            180                 185                 190
Lys Gln Val Ala Val Ala Glu Leu Leu Glu Asn Val Gly Gln Val Asn
        195                 200                 205
Glu His Asp Gly Gly Ala Gln Pro Gly Pro Val Pro Lys Ser Leu Gln
    210                 215                 220
Lys Gln Arg Arg Met Leu Glu Arg Leu Val Ser Ser Glu Ser Lys Asn
225                 230                 235                 240
Tyr Ile Ala Leu Asp Asp Phe Val Glu Ile Thr Lys Lys Tyr Ala Lys
                245                 250                 255
Gly Val Ile Pro Ser Ser Leu Phe Leu Gln Asp Asp Glu Asp Asp Asp
            260                 265                 270
Glu Leu Ala Gly Lys Ser Pro Glu Asp Leu Pro Leu Arg Leu Lys Val
        275                 280                 285
Val Lys Tyr Pro Leu His Ala Ile Met Glu Ile Lys Glu Tyr Leu Ile
    290                 295                 300
Asp Met Ala Ser Arg Ala Gly Met His Trp Leu Ser Thr Ile Ile Pro
305                 310                 315                 320
Thr His His Ile Asn Ala Leu Ile Phe Phe Phe Ile Ile Ser Asn Leu
                325                 330                 335
Thr Ile Asp Phe Phe Ala Phe Phe Ile Pro Leu Val Ile Phe Tyr Leu
            340                 345                 350
Ser Phe Ile Ser Met Val Ile Cys Thr Leu Lys Val Phe Gln Asp Ser
        355                 360                 365
Lys Ala Trp Glu Asn Phe Arg Thr Leu Thr Asp Leu Leu Leu Arg Phe
    370                 375                 380
Glu Pro Asn Leu Asp Val Glu Gln Ala Glu Val Asn Phe Gly Trp Asn
385                 390                 395                 400
His Leu Glu Pro Tyr Ala His Phe Leu Leu Ser Val Phe Phe Val Ile
                405                 410                 415
Phe Ser Phe Pro Ile Ala Ser Lys Asp Cys Ile Pro Cys Ser Glu Leu
            420                 425                 430
Ala Val Ile Thr Gly Phe Phe Thr Val Thr Ser Tyr Leu Ser Leu Ser
        435                 440                 445
Thr His Ala Glu Pro Tyr Thr Arg Arg Ala Leu Ala Thr Glu Val Thr
    450                 455                 460
Ala Gly Leu Leu Ser Leu Leu Pro Ser Met Pro Leu Asn Trp Pro Tyr
465                 470                 475                 480
Leu Lys Val Leu Gly Gln Thr Phe Ile Thr Val Pro Val Gly His Leu
                485                 490                 495
Val Val Leu Asn Val Ser Val Pro Cys Leu Leu Tyr Val Tyr Leu Leu
            500                 505                 510
Tyr Leu Phe Phe Arg Met Ala Gln Leu Arg Asn Phe Lys Gly Thr Tyr
        515                 520                 525
Cys Tyr Leu Val Pro Tyr Leu Val Cys Phe Met Trp Cys Glu Leu Ser
```

```
    530                 535                 540
Val Val Ile Leu Leu Glu Ser Thr Gly Leu Gly Leu Leu Arg Ala Ser
545                 550                 555                 560
Ile Gly Tyr Phe Leu Phe Leu Phe Ala Leu Pro Ile Leu Val Ala Gly
                565                 570                 575
Leu Ala Leu Val Gly Val Leu Gln Phe Ala Arg Trp Phe Thr Ser Leu
            580                 585                 590
Glu Leu Thr Lys Ile Ala Val Thr Val Ala Val Cys Ser Val Pro Leu
        595                 600                 605
Leu Leu His Trp Trp Thr Lys Ala Ser Phe Ser Val Val Gly Met Val
    610                 615                 620
Lys Ser Leu Thr Arg Ser Ser Met Val Lys Leu Ile Leu Val Trp Leu
625                 630                 635                 640
Thr Ala Ile Val Leu Phe Cys Trp Phe Tyr Val Tyr Arg Ser Glu Gly
                645                 650                 655
Met Lys Val Tyr Asn Ser Thr Leu Thr Trp Gln Gln Tyr Gly Ala Leu
            660                 665                 670
Cys Gly Pro Arg Ala Trp Lys Glu Thr Asn Met Ala Arg Thr Gln Ile
        675                 680                 685
Leu Cys Ser His Leu Glu Gly His Arg Val Thr Trp Thr Gly Arg Phe
    690                 695                 700
Lys Tyr Val Arg Val Thr Asp Ile Asp Asn Ser Ala Glu Ser Ala Ile
705                 710                 715                 720
Asn Met Leu Pro Phe Phe Ile Gly Asp Trp Met Arg Cys Leu Tyr Gly
                725                 730                 735
Glu Ala Tyr Pro Ala Cys Ser Pro Gly Asn Thr Ser Thr Ala Glu Glu
            740                 745                 750
Glu Leu Cys Arg Leu Lys Leu Leu Ala Lys His Pro Cys His Ile Lys
        755                 760                 765
Lys Phe Asp Arg Tyr Lys Phe Glu Ile Thr Val Gly Met Pro Phe Ser
    770                 775                 780
Ser Gly Ala Asp Gly Ser Arg Ser Arg Glu Glu Asp Asp Val Thr Lys
785                 790                 795                 800
Asp Ile Val Leu Arg Ala Ser Ser Glu Phe Lys Ser Val Leu Leu Ser
                805                 810                 815
Leu Arg Gln Gly Ser Leu Ile Glu Phe Ser Thr Ile Leu Glu Gly Arg
            820                 825                 830
Leu Gly Ser Lys Trp Pro Val Phe Glu Leu Lys Ala Ile Ser Cys Leu
        835                 840                 845
Asn Cys Met Ala Gln Leu Ser Pro Thr Arg Arg His Val Lys Ile Glu
    850                 855                 860
His Asp Trp Arg Ser Thr Val His Gly Ala Val Lys Phe Ala Phe Asp
865                 870                 875                 880
Phe Phe Phe Phe Pro Phe Leu Ser Ala Ala
                885                 890
```

<210> 5
<211> 890
<212> PRT
<213> Mus musculus

<400> 5

```
Met Asn Ser Gly Thr Pro Pro Pro Ser Pro Ser Gly Pro Pro Pro Pro
 1               5                  10                  15
Pro Ala Pro Gln Pro Gln Ala Arg Ala Arg Leu Asn Ala Thr Ala Ser
            20                  25                  30
Leu Glu Gln Asp Lys Ile Glu Pro Pro Arg Ala Pro Arg Pro Gln Ala
        35                  40                  45
Asp Pro Ser Ala Gly Arg Ser Ala Gly Glu Ala Ala Ala Pro Glu Pro
```

```
    50                  55                  60
Arg Ala Pro Gln Thr Gly Ser Arg Glu Glu Thr Asp Arg Ala Gly Pro
65                  70                  75                  80
Met Lys Ala Asp Val Glu Ile Pro Phe Glu Glu Val Leu Glu Lys Ala
                85                  90                  95
Lys Ala Gly Asp Pro Lys Ala Gln Thr Glu Val Gly Lys His Tyr Leu
            100                 105                 110
Arg Leu Ala Asn Asp Ala Asp Glu Glu Leu Asn Ser Cys Ser Ala Val
        115                 120                 125
Ala Trp Leu Ile Leu Ala Ala Lys Gln Gly Arg Arg Glu Ala Val Lys
    130                 135                 140
Leu Leu Arg Arg Cys Leu Ala Asp Arg Lys Gly Ile Thr Ser Glu Asn
145                 150                 155                 160
Glu Ala Glu Val Lys Gln Leu Ser Ser Glu Thr Asp Leu Glu Arg Ala
                165                 170                 175
Val Arg Lys Ala Ala Leu Val Met Tyr Trp Lys Leu Asn Pro Lys Lys
            180                 185                 190
Lys Lys Gln Val Ala Val Ser Glu Leu Leu Glu Asn Val Gly Gln Val
        195                 200                 205
Asn Glu Gln Asp Gly Gly Val Gln Pro Gly Pro Val Pro Lys Ser Leu
    210                 215                 220
Gln Lys Gln Arg Arg Met Leu Glu Arg Leu Val Ser Ser Glu Ser Lys
225                 230                 235                 240
Asn Tyr Ile Ala Leu Asp Asp Phe Val Glu Leu Thr Lys Lys Tyr Ala
                245                 250                 255
Lys Gly Ile Ile Pro Thr Asn Leu Phe Leu Gln Asp Glu Asp Glu Asp
            260                 265                 270
Glu Asp Glu Leu Ala Gly Lys Ser Pro Glu Asp Leu Pro Leu Arg Gln
        275                 280                 285
Lys Val Val Lys Tyr Pro Leu His Ala Ile Met Glu Ile Lys Glu Tyr
    290                 295                 300
Leu Ile Asp Val Ala Ser Lys Ala Gly Met His Trp Leu Ser Thr Ile
305                 310                 315                 320
Val Pro Thr His His Ile Asn Ala Leu Ile Phe Phe Phe Ile Ile Ser
                325                 330                 335
Asn Leu Thr Ile Asp Phe Phe Ala Phe Phe Ile Pro Leu Val Val Phe
            340                 345                 350
Tyr Leu Ser Phe Val Ser Met Val Ile Cys Thr Leu Lys Val Phe Gln
        355                 360                 365
Asp Ser Lys Ala Trp Glu Asn Phe Arg Thr Leu Thr Asp Leu Leu Leu
    370                 375                 380
Arg Phe Glu Pro Asn Leu Asp Val Glu Gln Ala Glu Val Asn Phe Gly
385                 390                 395                 400
Trp Asn His Leu Glu Pro Tyr Ile His Phe Leu Leu Ser Val Val Phe
                405                 410                 415
Val Ile Phe Ser Phe Pro Leu Ala Ser Lys Asp Cys Ile Pro Cys Ser
            420                 425                 430
Glu Leu Ala Val Ile Ser Thr Phe Phe Thr Val Thr Ser Tyr Met Ser
        435                 440                 445
Leu Ser Ser Ser Ala Glu Pro Tyr Thr Arg Arg Ala Leu Val Thr Glu
    450                 455                 460
Val Ala Ala Gly Leu Leu Ser Leu Leu Pro Thr Val Pro Val Asp Trp
465                 470                 475                 480
Arg Phe Leu Lys Val Leu Gly Gln Thr Phe Phe Thr Val Pro Val Gly
                485                 490                 495
His Phe Ile Ile Leu Asn Val Ser Leu Pro Cys Leu Leu Tyr Val Tyr
            500                 505                 510
Leu Phe Tyr Leu Phe Phe Arg Met Ala Gln Leu Arg Asn Phe Lys Gly
        515                 520                 525
```

```
Thr Tyr Cys Tyr Leu Val Pro Tyr Leu Val Cys Phe Met Trp Cys Glu
    530                 535                 540
Leu Ser Val Val Ile Leu Leu Gln Ser Thr Gly Leu Gly Leu Val Arg
545                 550                 555                 560
Ala Ser Ile Gly Tyr Phe Leu Phe Leu Phe Ala Leu Pro Ile Leu Val
                565                 570                 575
Ala Gly Leu Ala Leu Met Gly Thr Val Gln Phe Ala Arg Trp Phe Leu
            580                 585                 590
Ser Leu Asp Leu Thr Lys Ile Met Val Thr Thr Val Ile Cys Gly Val
        595                 600                 605
Pro Leu Leu Phe Arg Trp Trp Thr Lys Ala Asn Phe Ser Val Met Gly
    610                 615                 620
Met Val Lys Ser Leu Thr Lys Ser Ser Met Val Lys Leu Ile Leu Val
625                 630                 635                 640
Trp Leu Thr Ala Ile Leu Leu Phe Cys Trp Phe Tyr Val Tyr Arg Ser
                645                 650                 655
Glu Gly Met Lys Val Tyr Asn Ser Thr Leu Thr Trp Gln Gln Tyr Gly
            660                 665                 670
Phe Leu Cys Gly Pro Arg Ala Trp Lys Glu Thr Asn Met Ala Arg Thr
        675                 680                 685
Gln Ile Leu Cys Ser His Leu Glu Gly His Arg Val Thr Trp Thr Gly
    690                 695                 700
Arg Phe Lys Tyr Val Arg Val Thr Glu Ile Asp Asn Ser Ala Glu Ser
705                 710                 715                 720
Ala Ile Asn Met Leu Pro Phe Phe Leu Gly Asp Trp Met Arg Cys Leu
                725                 730                 735
Tyr Gly Glu Ala Tyr Pro Ser Cys Ser Ser Gly Asn Thr Ser Thr Ala
            740                 745                 750
Glu Glu Glu Leu Cys Arg Leu Lys Gln Leu Ala Lys His Pro Cys His
        755                 760                 765
Ile Lys Lys Phe Asp Arg Tyr Lys Phe Glu Ile Thr Val Gly Met Pro
    770                 775                 780
Phe Gly Thr Asn Gly Asn Arg Gly His Glu Glu Asp Asp Ile Thr Lys
785                 790                 795                 800
Asp Ile Val Leu Arg Ala Ser Ser Glu Phe Lys Asp Val Leu Leu Asn
                805                 810                 815
Leu Arg Gln Gly Ser Leu Ile Glu Phe Ser Thr Ile Leu Glu Gly Arg
            820                 825                 830
Leu Gly Ser Lys Trp Pro Val Phe Glu Leu Lys Ala Ile Ser Cys Leu
        835                 840                 845
Asn Cys Met Thr Gln Leu Ser Pro Ala Arg Arg His Val Lys Ile Glu
    850                 855                 860
Gln Asp Trp Arg Ser Thr Val His Gly Ala Leu Lys Phe Ala Phe Asp
865                 870                 875                 880
Phe Phe Phe Phe Pro Phe Leu Ser Ala Ala
                885                 890
```

<210> 6
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 6
tgtaaaacga cggccagtct cgtgcagaag gccgcgct 38

<210> 7

<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 7
caggaaacag ctatgaccgc ccacagccac cgcgccac 38

<210> 8
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 8
tgtaaaacga cggccagtct gtctccagca gacactaa 38

<210> 9
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 9
caggaaacag ctatgaccca caatgctgaa ctgcagag 38

<210> 10
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 10
tgtaaaacga cggccagtct gaagaccctc atgccttg 38

<210> 11
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 11
caggaaacag ctatgaccac acttctctgt gggctgtg 38

<210> 12
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 12
tgtaaaacga cggccagttc ggagaatctg gaggctga 38

<210> 13
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 13
caggaaacag ctatgaccca ttacaagctg ctcaaccc 38

<210> 14
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 14
tgtaaaacga cggccagtcg aaagccttcc aggcagag 38

<210> 15
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 15
caggaaacag ctatgaccct atgggaaggt cctggctc 38

<210> 16
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 16
tgtaaaacga cggccagtct aggaacagtg cgccagtt 38

<210> 17
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 17

caggaaacag ctatgaccat ggagtcgcac aggaagga 38

<210> 18
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 18
tgtaaaacga cggccagtgc ccatgctgtt ttctctca 38

<210> 19
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 19
caggaaacag ctatgacccc gaggacacat ccttatga 38

<210> 20
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 20
tgtaaaacga cggccagtcc tcgttcccac gtaccatc 38

<210> 21
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 21
caggaaacag ctatgaccgt agcagtaggt gcccttga 38

<210> 22
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 22
tgtaaaacga cggccagtcc tggtcgtcct caatgtca 38

<210> 23
<211> 38

<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 23
caggaaacag ctatgaccca tagaaccagc agaacagc 38

<210> 24
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 24
tgtaaaacga cggccagttg gttcacgtct ctggagct 38

<210> 25
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 25
caggaaacag ctatgaccga gttgtagacc ttcatgcc 38

<210> 26
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 26
tgtaaaacga cggccagtgg gcatgaaggt ctacaact 38

<210> 27
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 27
caggaaacag ctatgaccga acttcttgat gtggcagg 38

<210> 28
<211> 38
<212> DNA
<213> Artificial Sequence

<220>

<223> /note= "synthetic construct - primer"

<400> 28
tgtaaaacga cggccagtct ggatgcgctg cctctacg 38

<210> 29
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 29
caggaaacag ctatgacctc aggccgccga caggaatg 38

<210> 30
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 30
tgtaaaacga cggccagttc gccttcgact tctttttc 38

<210> 31
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 31
caggaaacag ctatgacccc aaacaaataa gaaatgct 38

<210> 32
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 32
gcccagctct cgcccaccag 20

<210> 33
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 33
tcaggccgcc gacaggaatg 20

<210> 34
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 34
cctggtcgtc ctcaatgtca 20

<210> 35
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 35
ggtagggcac aaggtagcag 20

<210> 36
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 36
gggcatgaag gtctacaact cca 23

<210> 37
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 37
ccgtagaggc agcgcatcca gtcgccgacc tagaac 36

<210> 38
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 38
gagggcatga aggtctacaa 20

<210> 39
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 39
cccacggtaa tctcaaactt 20

<210> 40
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 40
tagtgtgccc ctgctgttgc 20

<210> 41
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 41
cccacggtaa tctcaaactt 20

<210> 42
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 42
cggtttcgga gcaacttcgc 20

<210> 43
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> /note= "synthetic construct - primer"

<400> 43
cacctcagcc tcgttctcag 20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/42429

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07H 21/04; C07K 14/46, 16/28; C12Q 1/68; C12P 19/34
US CL :536/23.5, 24.31; 530/350, 387.9; 435/6, 91.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5, 24.31; 530/350, 387.9; 435/6, 91.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE
search terms:wolfram syndrome

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T,E	HARDY et al. Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in WFS1. American Journal of Human Genetics. November 1999, Vol. 65, No. 5, pages 1279-1290.	1-17
A	COLLIER et al. Linkage of Wolfram syndrome to chromosome 4p16.1 and evidence for heterogeneity. American Journal of Human Genetics. October 1996, Vol. 59, No. 4, pages 855-863.	1-17

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
29 JANUARY 2000	11 FEB 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DAVID S. ROMEO Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/22429

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	STROM et al. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. Human Molecular Genetics. December 1998, Vol. 7, No. 13, pages 2021-2028, see entire document.	1-17
P, X	INOUE et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nature Genetics, October 1998, Vol. 20, No. 2, pages 143-148, see entire document.	1-17
X	US 5,578,444 A (EDWARDS et al.) 26 November 1996, columns 115-118, SEQ ID NO:36.	6-8